THE ROLE OF PRINCIPLES IN INSTRUCTIONS FOR PROCEDURAL TASKS: TIMING OF USE, METHOD OF STUDY, AND PROCEDURAL INSTRUCTION SPECIFICITY

A Dissertation
Presented to
The Academic Faculty

by

Elsa Eiriksdottir

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in Engineering Psychology

Georgia Institute of Technology

December 2011

THE ROLE OF PRINCIPLES IN INSTRUCTIONS FOR PROCEDURAL TASKS: TIMING OF USE, METHOD OF STUDY, AND PROCEDURAL INSTRUCTION SPECIFICITY

Approved by:

Dr. Richard Catrambone, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Wendy Rogers
School of Psychology
Georgia Institute of Technology

Dr. Frank Durso
School of Psychology
Georgia Institute of Technology

Dr. Mark Guzdial
College of Computing
Georgia Institute of Technology

Dr. Donna Llewellyn
Center for Enhancement of Teaching and Learning
Georgia Institute of Technology

Date Approved: October 12, 2011

I dedicate this work to my family. Particularly to my amazing husband, Unnar, who has encouraged me every step of the way and our daughter Vaka, who brings light and joy to the mundane and helps me keep things in perspective. I also want to thank my parents for their unwavering loving support and never allowing me to falter in the belief that I could do anything I set my mind to.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Richard Catrambone for his guidance, support, patience, and endless editing. I would also like to thank my committee members Dr. Wendy Rogers, Dr. Mark Guzdial, Dr. Frank Durso, and Dr. Donna Llewellyn for their helpful suggestions and encouraging criticism. Finally, I want to thank all of the people who helped me so tirelessly collect and code the data: Merilee Berry, Chance Hughes, Kaitlyn Kreager, Joe Dagosta, Hannah Krimm, and Kevin Strika.

A special thanks and appreciation goes to Warren Rhude of Simutech Multimedia for kindly allowing me to use the Troubleshooting Electrical Circuits software in my research and graciously providing personal support throughout the process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Including domain rules and generalities (principles) in instructions for procedural tasks is believed to help learners understand the task domain (or the system), and in turn make them better able to complete tasks. However, equivocal results of prior research indicate that principles are not always beneficial. The goal of the current research was to delineate the characteristics of the conditions under which principles are useful. In two studies I investigated the impact of the timing of principle use, the method used to study the principles, and the specificity of the procedural instructions accompanying the principles. The first study showed that the timing of principle use (studying the principles before, during, or after completing training tasks) did not affect declarative (knowledge of the system) or procedural learning (troubleshooting task performance). Therefore, the commonly advocated idea that principles should be provided before task engagement was not supported. Neither was the hypothesis that using principles while solving tasks would enhance procedural learning. When learners summarized the principles, they demonstrated better declarative learning compared to when they just read the principles. Better declarative learning was associated with better procedural learning, but the relationship between understanding and using a system is likely not as direct as often assumed. In the second study declarative and procedural learning were enhanced when the principles were accompanied by general rather than detailed procedural instructions. General procedural instructions appeared to encourage task engagement and the effective use of principles although this effect was reduced if leaners were required to summarize the principles rather than simply read them. Together the results of the two studies reveal

how the learning situation and instructional materials can be constructed to create conditions where principles enhance learning and subsequent performance.

INTRODUCTION

People commonly find themselves in a situation where they have to learn how to do something new, such as use unfamiliar software, learn a new craft, prepare a novel dish, or apply a new method in math. In these situations people commonly use instructions to understand what they need to do and how the domain or system works. Providing instructions is a common and expected method of communicating with someone learning a procedural task. Instructions are messages that provide information about the system, the domain, the product, or the task. Defining instructions as messages makes clear that instructions represent information provided by an authority in the domain, whether it is the designer of the system or a domain expert, and that they are intended to be received by the learner or domain novice (Lannon, 2008; Mayer, 1981). Instructions can be found in various situations: They are usually provided with most consumer products (e.g., a microwave, software, or digital camera), they are included in textbooks (e.g., on programming, physics, or statistics), and instructional materials for skill based tasks (e.g., cooking, knitting, or carpentry). Most commonly, instructions are written text, sometimes with accompanying diagrams or pictures. In all cases the goal of the instructions is to convey how to complete tasks within a domain by describing the procedure involved and by providing information that might make performance and learning easier.

There are different ways of distinguishing between various types of instructions, but a common distinction is between procedural and declarative instructions (Karreman, 2004; Karreman, Ummelen, & Steehouder, 2005). Procedural instructions are what most people refer to when they speak of instructions. They are stepwise descriptions of how to carry out a task and guide people by describing the conditions for carrying out a step of the procedure, the actions required (sequence of movements and tool manipulations), and how the states of the system change as a result of these actions. Procedural instructions are most often organized as a series of successive steps that need to be carried out to complete the task (Andrews, 2001; Bibby & Payne, 1993; Farkas, 1999; Guthrie, Bennett, & Weber, 1991; Karreman, et al., 2005; Konoske & Ellis, 1991; Lannon, 2008). For example, procedural instructions for using a self-service gas pump might describe following a series of steps: 1) using a card or cash to pay for the gas and how one knows the pump is ready, 2) how to pull the handle of the gas pump to pump the gas, and how the dollars and gallon counters change as a result, and 3) how to end the transaction.

Declarative instructions are all other types of information that are not procedural and are intended to aid users in figuring out what to do to achieve their goal (Karreman, et al., 2005). This can include information about the interface, how to make optimal use of the system or device, and how the system works. Information about how the system works (principles) is the most common type of declarative instructions, and in the literature has been variously referred to as principles (Bibby & Payne, 1993; Catrambone, 1990, 1995; Karreman, et al., 2005), functional information (Smith & Goodman, 1984), supportive information (Kester, Kirschner, Van Merrienboer, & Baumer, 2001), system information (Karreman & Steehouder, 2004), and models (Mayer, 1989). Principles contain information about the internal workings of the system or device, and are therefore not directly task oriented, but instead describe the logic of the system and the cause and effect mechanisms that determine outcome when some variable is manipulated. That is, principles are domain rules and generalities, and are often described as "how it works" information, to be contrasted with "how to do it" information that is characteristic of procedural information (Bibby & Payne, 1993; Duff & Barnard, 1990).

As the definition of principles indicates, principles are relevant where the domain involves a system of some sort. A system can be defined as a coherent collection of parts that interact and variables that make up the system have a causal relationship (Mayer, 1989; Simon, 1969). Example of systems can be devices such as cameras or bicycles, computer software and games, domains such as mathematics or programming, and physical systems such as chemical plants or distilleries. In each case, there are a set of principles that describe how elements within the system interact and affect each other, and subsequently provide information about some rules or generalities in the system.

Researchers have hypothesized that providing principles in instructions is helpful for learners because it will provide them with a better understanding of the system they are using and that in turn will aid learning how to use the system. Research however has not yielded unequivocal results and there is reason to believe that other factors in the learning situation might determine whether principles are helpful. The goal of the current research is to investigate the conditions that determine whether providing principles in instructions was beneficial for declarative learning (reflected in the knowledge of the system) and procedural learning (reflected in the ability to complete tasks).

Why Provide Principles?

The rationale for providing principles in instructions is based on the premise that it increases the understanding that users have of the system and having better understanding facilitates learning how to use the system (procedural learning). Having a more comprehensive conceptual understanding is believed to be especially important when the learner has to figure out how to do new tasks or deal with unexpected situations (Bibby & Payne, 1993; Gott, Lajoie, & Lesgold, 1991; Karreman & Steehouder, 2004; Mayer, 1981). This line of reasoning therefore predicts that providing principles in instructions will be particularly helpful in situations that require problem solving (e.g., in troubleshooting) because understanding how the system works will make the users better able to reason about the system and infer the steps needed to be taken to complete unfamiliar tasks (Kieras & Bovair, 1984).

In the research literature on instructions, the benefit of principles is generally attributed to the mental representation that the user builds from the instructions and from interacting with the system. Most commonly these mental representations are referred to as mental models, and the general hypothesis is that principles allow the users to build a more comprehensive and coherent mental model; making users better able to understand and predict the effect of interacting with the system (Bibby & Payne, 1993; Borgman, 1999; Duff & Barnard, 1990; Karreman, et al., 2005; Kieras & Bovair, 1984; Patrick & Haines, 1988).

A mental model of a system is the users' conceptual knowledge of the system and represents the structure and relationships of the elements within it. Mental models are also a mechanism for making inferences and testing hypotheses about the system (Borgman, 1999; Norman, 1987). The exact conceptualizations of mental models can vary considerably between researchers, but mental models are usually conceptualized as being dynamic, developing with experience, and being updated with current information (Bibby & Payne, 1993; Borgman, 1999; Norman, 1987).

The hypothesis that principles lead to a better mental model and enhanced transfer includes two separate but connected assumptions. First, that learning from principles leads the learner to develop to a better mental model of the system. Researchers often do not agree on just how complete mental models need to be or how exactly new experience changes them (Waern, 1993), but generally one can assume that a better understanding of how elements in the system interact would constitute a better mental model. The second assumption inherent in the mental model account is that having a better mental model will lead to better procedural learning, especially when the learner has to complete unfamiliar or new tasks within the system, such as involving troubleshooting or transfer of learning.

Therefore, according to advocates of the mental model account it can be generally hypothesized that studying the principles of a system leads to a more comprehensive mental representation or knowledge of the system (because the learning is organized according to rules governing the system), and this in turn leads to better performance of tasks within the system, especially when the tasks are new or involve problem solving (because a good mental representation or knowledge framework allows users to correctly infer how to complete new tasks with the system). However, the evidence from research literature on using principles in instructions for procedural tasks has not provided clear support for this account.

Do Principles Enhance Learning?

A number of studies have aimed at demonstrating the benefit of providing principles in instructions, but the general results have been inconclusive. Some studies have shown that providing principles leads to better learning and transfer (Borgman, 1999; Catrambone, 1995; Karreman & Steehouder, 2003; Karreman & Steehouder, 2004; Kieras & Bovair, 1984); others have demonstrated some benefit only for transfer (Kontogiannis & Sheperd, 1999; Patrick & Haines, 1988; Smith & Goodman, 1984); and some have shown no effect at all (Berry & Broadbent, 1984; Morris & Rouse, 1985; Reder, Charney, & Morgan, 1986).

To date there has been no clear explanation for this disparity in findings and comparison between studies is difficult because of the different methodologies and tasks used in the studies (Karreman, 2004). However, the diversity of methods used can also be taken to indicate that whether principles are helpful depends on factors in the learning situation that have not been identified properly yet. Three general factors that seem to vary in the literature and could be influential are: (1) how the principles are studied, (2) when the principles are used, and (3) what kind of instructional information is provided, both in terms of principles and procedural instructions.

How the principles are used by the learner is important because research has clearly shown that different levels and ways of processing instructional information can lead to very different learning outcomes (Chi, 2009; Chi, de Leeuw, Chiu, & LaVancher, 1994; Kolodner, Owensby, & Guzdial, 2004; Pirolli & Recker, 1994; Wittwer & Renkl, 2008). Therefore comparing groups of participants that either receive principles or not can be problematic if it is not clear how the participants are using the principles. For example, if the learners in a group that receives principles to help them learn do not study the principles, comparison with a group that does not receive the principles should not yield any difference. No difference could mean either that the principles are not effective or that they were not studied. Therefore, it is important to consider the method used to study the principles.

When the principles are used refers to the timing of use. Most commonly, this refers to whether the principles are studied before or while the learner interacts with the system and carries out tasks. Instructions are typically designed from the standpoint that they should be used before attempting to complete tasks and research on advance organizers suggests that there might be benefits to doing so (Ausubel, 1960; Ganier, 2004; Mayer, 2003; Mayer & Bromage, 1980). Other researchers have emphasized that instructional information is most beneficial when studied in the context of doing the task (Alterman, Zito-Wolf, & Carpenter, 1991; Carroll, 1990, 1998; Kester, Kirschner, & Van Merrienboer, 2006; Kester, et al., 2001).

The kind of instructional information provided refers to how principles and procedural instructions are defined and implemented. Research has indicated that principles are beneficial insofar as they are relevant to the task at hand and relate the controls to the structure and function of the system (Kieras & Bovair, 1984). It is therefore important to consider carefully the kind of principles that are included in the instructions and the extent to which they relate to the task that the learner is trying to accomplish. The definition and implementation of procedural instructions is also important because research has indicated that the specificity and amount of procedural instructions determines whether principles are helpful (Catrambone, 1995; Duff & Barnard, 1990).

Each of these three factors can affect whether providing principles in instructions will help procedural learning and transfer. Each will be considered in turn.

How Are Principles Used?

A typical study on the effects of principles compares two groups of learners, one that receive only procedural instructions, and one that also receives principles (Karreman & Steehouder, 2003; Karreman & Steehouder, 2004; Kieras & Bovair, 1984; Konoske & Ellis, 1991; Morris & Rouse, 1985; Patrick & Haines, 1988; Payne, Howes, & Hill, 1992). The comparison between the two groups on some learning outcome is then assumed to indicate the usefulness of the principles. This assumption is problematic because there is no guarantee that the participants actually used the information provided, let alone learned from it. In the research literature, a variety of methods of using the principles can be seen: Some researchers have tried to guarantee learning by training to criterion (e.g., Kieras & Bovair, 1984 who found benefits for procedural learning under specific circumstances), others only provide limited opportunity for study by having the principles read to the participants once (e.g., Berry & Broadbent, 1984 who found benefits for declarative but not procedural performance), and yet others neither measure nor control how the principles are used (e.g., Karreman, 2004; Karreman & Steehouder, 2003).

The differences in methods make it difficult to compare studies, and to understand under what circumstances principles enhance declarative or procedural learning and when non-significant effects are due to principles not being used. Aside from pure methodological considerations, it is important to specify the method and context of studying the principles in instructions because these have been shown to determine the learning that takes place, both declarative and procedural.

Engagement and Active Processing

A few related theoretical accounts emphasize active cognitive processing of the instructional material and learning experience to be the important method for fostering both declarative and procedural learning. This active engagement in the learning situation requires the learners to engage in effortful cognitive processing, for example by self-explanations, relating what is being learned to prior knowledge, or by doing a task that is important or meaningful to the learner (Ausubel, 1960; Carroll, 1998; Chi, 2009; Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Davis & Wiedenbeck, 1998; Wiedenbeck & Zila, 1997).

The minimalist approach to instructions emphasizes active engagement through learners completing tasks that are meaningful to them. A central assumption is that the most effective mechanism for learning is to get learners to actively construct their own knowledge through the experience of doing the task (Carroll, 1987; Carroll & Mack, 1987; Carroll, Mack, Lewis, Grischkowsky, & Robertson, 1985). This idea of active processing for learning is also represented in learning-by-doing, discovery-learning, and constructivist accounts of learning (Carroll, 1998; Kolodner, et al., 2004; Quintana, Carra, Krajcik, & Soloway, 2002; Van Der Meij, 2003; Van Der Meij & Lazonder, 1993). In these approaches an important mechanism for learning is to have the learners engage with and complete tasks.

Other researchers have, instead of emphasizing task engagement, emphasized cognitive engagement through self-explanations and deliberate practice. Chi and colleagues (Chi, 2009; Chi, et al., 1989; Chi, et al., 1994) have stressed the role of self-explanations in improving understanding and determining level of learning and Ericsson and colleagues have focused on deliberate practice where activities are specifically structured to improve performance (Ericsson, Krampe, & Tesch-Romer, 1993). What these approaches have in common is that learners have to monitor and reflect on their own knowledge and performance.

In an early study, Chi et al. (1989), found that quality of self-explanations the learners spontaneously provided was different for good and poor learners. The good learners (more successful at problem solving) explained and justified each action while studying example exercises and tried to relate what they were doing to principles and concepts provided in the instructional materials. The poor learners seldom tried to self-explain the example exercises, and had difficulty connecting what they were doing to the information provided in the instructional materials. The researchers concluded that elaboration during exercises and while using the instructional materials was an important strategy for good learning.

Even if self-explanations were found to be characteristics of good learners, later research has shown that self-explanations can be taught and encouraged through different study activities. In particular, Chi (2009) proposed a framework describing different categories of activities learners can engage in to enhance learning. According to Chi, engaging in any overt activities of engagement (e.g., summarizing the information) will lead to better learning when compared with engaging in passive activities (e.g., reading).

Chi et al. (1994) have proposed a few reasons why self-explanations are successful. First, self-explaining presumably allows learners to identify conflicting information and provides them with opportunities to correct false assumptions or erroneous conclusions. Second, by self-explaining, the learners attempt to relate the new knowledge to prior knowledge as shown by their results that 30% of the self-explanations were found to result from learners integrating new information with old.

In the same vein, Ericsson and colleagues have demonstrated that learners successful at improving their performance actively reflect on their performance while searching for new methods to improve it (Ericsson, et al., 1993). In addition, an important component of deliberate practice is to consider prior domain knowledge and often the successful strategy for improving performance relies upon taking advantage of this domain knowledge.

Overall, evidence supports the idea that having learners engage in active processing while using instructional materials leads to effective learning (Alterman, et al., 1991; Carroll, 1990; Chi, 2009; Chi, et al., 1994; Kolodner, et al., 2004; Quintana, et al., 2002). What exactly this active processing entails is not always agreed upon: Some emphasize the importance of providing meaningful task context where others emphasize the importance of cognitive activities. However, these differences are probably superficial because one can argue that providing a meaningful task is one way of encouraging learners to cognitively engage via problem solving.

Indeed, the consensus in the research literature seems to be that the important parts of active processing, involves both problem solving (hypothesis testing) and organization of information into a coherent representation, sometimes with the help of prior knowledge (assimilation) (Mayer, 2003; Mayer & Bromage, 1980; Wittwer & Renkl, 2008). These cognitive activities, problem solving and assimilation, presumably help learners construct their own knowledge from the learning activities.

Learning Processes

Problem Solving

Problem solving has been defined as taking place when someone has a goal but not a clear sense of how to reach that goal (Mayer, 1989). When users find themselves in this situation, instructions are likely to be helpful. Novick and Bassok (2005) made an interesting claim about problem solving – that thinking is required “whenever one cannot go from the given situation to the desired situation simply by action (i.e. by the performance of obvious operations)” (p. 321). This statement reflects a common idea in research on instructions; if learners are provided with too detailed or complete instructions they will simply follow the instructions without engaging cognitively with the task. This is detrimental to learning as learning requires a certain amount effortful cognitive activities (Catrambone & Yuasa, 2006; Chi, et al., 1989; Schmidt & Bjork, 1992; Wittwer & Renkl, 2008).

The evidence suggests that engaging in learning is an inherently effortful process; both Schmidt and Bjork (1992) and Schneider (1985) reviewed findings from research on learning and concluded that factors that lead to efficient initial performance seem often to lead to poor learning, and factors that make initial performance difficult or more effortful seem to lead to better learning. Schmidt and Bjork stated “the general point is that certain "difficult" training conditions may foster various kinds of processing activities that are required for effective retention performance” (p. 214). This idea dovetails with research findings that problem solving is one reason why engagement and active processing during training enhances learning.

Assimilation

The other learning process proposed to be an important part of engagement and active processing is assimilation of newly learned information to prior knowledge and experience. Mayer (1989) proposed that when learning from principles the learner is guided to select, organize, and integrate information. That is, principles lead learners to attend to information on the conceptual aspects of the system (e.g., components, actions, states, and causal relations), organize the information around coherent explanations (i.e., build internal connections), and integrate the information with existing relevant knowledge (i.e., build external connections).

Reigeluth (1983) distinguished between different categories of prior knowledge that could facilitate organization and integration of new information. One category represents knowledge within the domain in question and includes, among other types of prior knowledge, superordinate knowledge and experiential knowledge. Superordinate knowledge is used in this context to refer to an idea that is more inclusive than the idea to-be-learned. According to Reigeluth, principles are one example of superordinate knowledge and attempting to assimilate new information with principles should enhance learning because it helps acquisition and organization of knowledge. By comparing and contrasting the new information to the principle, acquisition is enhanced, and relating new information to prior information helps create stable memory traces for the knowledge. This idea is akin to the theory on advance organizers discussed later. Experiential knowledge refers to memory of specific instances and these can serve to enhance understanding and provide examples of how generalities are applied. Therefore, it is important for acquisition of memory to relate new information to instances.

Experiential knowledge or instances can also be organized in accordance to new superordinate information, and thus help organization of memory and increase memory traces.

Overall, there is reason to believe that for learning to take place the learner has to encounter some difficulty during the learning episode. But, the question of what kind of difficulty rises immediately. It is unlikely that the difficulty encountered when using poorly designed or conceptualized instructions would be considered constructive for learning. Instead, prior discussion indicates that difficulty in this context should refer to engaging in effortful cognitive processes such as problem solving and assimilation. Given that a certain type of difficulty is desirable in a learning situation, the question becomes how much difficulty should the learner encounter? Research on cognitive load indicates that there are limits to how much difficulty should be introduced.

Subjective Workload

The idea that learning requires cognitive effort immediately raises questions of how much effort is enough and when it becomes too much? This is the topic of research on subjective workload. Subjective workload refers to experiencing the exertion of mental effort and is built on the assumption that people have limited cognitive capacity and that some situations or tasks will exceed this capacity. This has negative consequences, such as diminished learning, worse performance, and negative emotions (e.g. feelings of frustration or incompetence) (Hicks & Wierwille, 1979; Jex, 1988; Rubio, Diaz, Martin, & Puente, 2004).

In the learning domain, the relationship between subjective workload and learning has been largely investigated in the context of cognitive load theory. According to the cognitive load theory, effective learning can take place if the learner has cognitive resources available for self-explanations and schema acquisition (Paas, Renkl, & Sweller, 2003; Renkl & Atkinson, 2003; Sweller, 1988; Sweller, Chandler, Tierney, & Cooper, 1990). Learning requires an optimal use of limited working memory capacity and the amount and type of cognitive load imposed by a learning situation determines learning. (Paas, Camp, & Rikers, 2001; Paas, et al., 2003; Renkl & Atkinson, 2003; Sweller, 1988; Sweller, et al., 1990; Van Merrienboer, Kirschner, & Kester, 2003). The more cognitive load that is required for irrelevant processing, less is available for learning, and conversely, if cognitive resources are devoted to relevant processing learning is enhanced.

In terms of cognitive load theory, the potential problem with including principles in instructions for procedural tasks is that they can easily add to the cognitive load of the task. Either by making the task more complicated or by adding what the learner might view as unnecessary information for completing the task. Therefore, adding principles to instructions for procedural tasks can lead to cognitive overload, where the learner gets overwhelmed with information and is not able to engage in learning enhancing processes such as schema acquisition and self-explanations. The potential for overload means that the benefit of adding principles to instructions for procedural tasks must be clear and subjective workload must be measured.

Conclusion: The Method of Studying the Principles

Generally, research on learning suggests that effective learning is the result of active cognitive processing that is effortful, and involves some level of problem solving and assimilation of information. The learners need to be faced with a certain level of difficulty and cognitive load in the learning situation or during training, but too much subjective workload is counterproductive. Therefore, learners seem to require a certain level of support, but at the same time should be required to invest cognitive effort.

From this it would seem important that when studying effects of some manipulation of the learning situation, such as the influence of different types of instructions, to know or predict what the learners are doing during training. Are they actively engaging in problem solving? Are they attempting to assimilate information? Are they taking advantage of the support provided or choosing to explore the system on their own and use trial-and-error learning?

Mayer (2001) made the point that

> …active learning occurs when a learner applies cognitive processes to incoming materials – processes that are intended to help the learner make sense of the material. The outcome of active cognitive processing is the construction of a coherent mental representation, so active learning can be viewed as a process of model building. A *mental model* (or *knowledge structure*) represents the key parts of the presented material and their relations. (p. 51)

Mayer emphasized that understanding an instructional message involves constructing a knowledge structure, and that the instructional materials should have a coherent structure to assist the learner in building the mental model. And conversely, if instructions lack guidance for how to structure the presented material, the learner might be overwhelmed in his or her efforts to build a mental model. This conclusion represents the mental model account, and supports the idea that providing principles in instructions should be helpful for learning as they assist learners in constructing a coherent knowledge structure.

Mayer is also a proponent of another approach of how the learning situation can be constructed to enhance learning: organizing information should be provided before learning takes place (advance organizer). The idea that the timing of when information is provided is important is often implied in research on principles but rarely explicitly stated or tested.

When Are Principles Used?

In some studies on the effect of providing principles in instructions the principles were only available before the participants engage with tasks in the domain (Borgman, 1999; Kieras & Bovair, 1984; Kontogiannis & Sheperd, 1999) in others they were available during task completion (Catrambone, 1995; Karreman & Steehouder, 2003; Karreman & Steehouder, 2004). There is reason to believe that the timing of principle use might have an effect on the learning that takes place.

Advance Organizers

One way in which principles can help a learner is to act as a type of advance organizers. Ausubel (1960) proposed that providing an “advance introduction of relevant subsuming concepts (organizers)” (p. 267) could facilitate learning and retention of verbal material. Based on a hierarchical model of memory where concepts are organized under subsuming categories, Ausubel’s advance organizers were expected to deliver a learning advantage through an organization of subsuming categories for the to-be-learned material.

Advance organizers act as organizational cues that provide a framework to help the learner retain information better. Mayer and Bromberg (1980) further defined advance organizers as “stimulus (usually a prose passage) that (a) is presented prior to learning and (b) contains a system for logically organizing the incoming information into a unified structure” (p. 211). Principles provide logical organization of incoming information because they provide domain rules that instances fall under and therefore principles could easily become advance organizers if they are presented before learning. Indeed, using advance organizers has been categorized as a type of deductive learning method where the rule is provided and examples of the rule follow (Mayer, 2003).

Additionally, Langan-Fox, Waycott and colleagues (Langan-Fox, Platania-Phung, & Waycott, 2006; Langan-Fox, Waycott, & Albert, 2000) have proposed that advance organizers might be very beneficial in instructions for use. They have supported this idea by pointing out that one of the enduring differences between novices and experts is the schemata of the domain: Experts use the schemata to organize their domain knowledge. This means that advance organizers could be helpful by providing novices with a way of building schemata of the domain.

This argument is strikingly similar to the argument presented in the mental model account: Organizing information (principles) helps the learner build a mental representation of the to-be-learned information that either is assembled faster (more efficiently) or is better (more comprehensive). However, the theory on advance organizers has an explicit time component – the information should be provided before learning of instances takes place. The importance of timing is often assumed in studies based on the mental model account as the principles are provided before the learning episode takes place, but this assumption is rarely if ever explicitly stated and there does not seem to be much overlap between research on using principles in instructions and research on advance organizers.

Research on advance organizers has indicated they provide a way of organizing new information and can help learning under specific conditions: Advance organizers are helpful when the learners have little or no prior knowledge of the domain, when transfer of knowledge is needed (as opposed to recall), and when the organizers provide concrete as opposed to abstract information (Mayer, 2003; Mayer & Bromage, 1980). It is logical that advance organizers should be less helpful for learners who already have knowledge about the domain as compared to novices who do not, because the advance organizers might be superfluous or even conflict with the current organization knowledge. The condition that organizers have to be concrete and not abstract can be seen to be analogous to the findings that principles have to be directly relevant to the information that is to be learned (Kieras & Bovair, 1984). That is, provide concrete information about how to use the system.

Most research on advance organizers has focused on verbal learning and not procedural learning, but a couple of studies have also found benefit of using advance organizers for problem solving and procedural learning. Mayer (2003) reported that a group of learners receiving advance organizers not only had better recall of information from the lesson, but also had higher problem solving scores within the domain (how radars work). Langan-Fox et al., (2006) found that providing learners with text based advance organizers resulted in better procedural knowledge (on using mobile networks) as compared to groups of learners that did not get advance organizers or graphical organizers.

Taken together, the research on advance organizers suggest there might be benefits for learning to provide principles describing how the system works before learners start engaging with tasks. This is also implicitly assumed in the idea of mental models and how prior knowledge can help learning. However, there is also reason to believe principles might be more advantageous when used in the context of doing procedural tasks. The just-in-time information presentation method emphasizes that for procedural learning the instructions should be used in the context of the task at hand as it gives the information meaning and lowers extrinsic cognitive load.

Just-In-Time Information

Just-in-time (JIT) information presentation is a method advocated by proponents of cognitive load theory. The idea in its simplest terms is that instructions should be presented exactly when the learner needs them in the process of doing the task, as opposed to reading them before starting the task (Kester, et al., 2006; Kester, et al., 2001; Van Merrienboer, et al., 2003). By presenting the procedural information just-in-time, proceduralizing the declarative information is facilitated because the relevant information is active in working memory when it is needed in practice. This lowers cognitive load by reducing temporal split attention which results from the learner having to mentally integrate the information on the status of the task with the relevant part of the instructions retrieved from memory. By eliminating temporal split attention, more effective learning can take place because extraneous cognitive load is reduced and schema construction and automation become easier (Kester, et al., 2001; Van Merrienboer, et al., 2003).

Another line of reasoning suggests that providing instructions in the context of a task could lower intrinsic load because the task provides a meaningful context for the information provided in the instructions. Alterman et al. (1991) pointed out that instructions are often difficult to comprehend outside of the task, as they refer to components, actions, and states of the system as it is being used. Therefore, attempting to comprehend instructions outside of the context of the task should impose higher intrinsic cognitive load than doing so while engaging in a task. In addition, Alterman et al. proposed in their FLOABN (For a Lack of A Better Name) model that as people use instructions for use in their everyday lives, they build understanding through interactions with the system and the task they are trying to complete. Understanding in this context refers to the learner constructing a mental representation that corresponds to the situation, is coherent with information in semantic memory, and is interpreted in the context of the person's current goals. Here the important part of understanding is allowing the learner to generate actions from his or her representation.

The theoretical accounts discussed here do not specify exactly what kind of instructions they refer to, and one can assume that they are generally referring to procedural instructions. The rationale they provide can however easily apply to principles. Principles describe how elements of a system interact and this cause-and-effect relationship should presumably be easier to understand in the context of a task where these relationships become instantly apparent in the behavior of the system.

Conclusion: The Timing of Principle Use

The preceding discussion indicates that there are reasons to believe that the timing of principle use could affect learning. The research on advance organizers indicates that when learners study principles before engaging in training, their understanding of the system will be better and the knowledge they construct more coherent. However, both the just-in-time and the FLOABN model provide reasons to believe that providing instructional information in the context of training tasks can have a beneficial effect for learning, even if empirical support has been lacking so far.

In both cases the argument is made that learning will be better when people are provided with a relevant context for the to-be-learned information. This is analogous to the findings of Bransford and Johnson (1972) where the comprehension of information was dependent upon contextual knowledge. In their study, participants demonstrated better comprehension and recall when they had been provided with contextual information than if they had not. Bransford and Johnson stated that not just prior knowledge plays a role, but also that "certain information might be necessary for meaningful processing of the information in the first place" (p. 718). The question is whether principles provide a context for tasks, or whether the tasks provide a context for the principles.

What Kind of Principles and Procedural Instructions?

Research on instructions has indicated that whether including principles in instructions can help learning might depend both on the principles themselves and the specificity of the procedural instructions provided.

Relevance of Principles

In prior research on whether adding principles to instructions is helpful for learning, there is generally a consistent definition of principles as instructions that describe the internal workings of a system (Catrambone, 1995; Karreman & Steehouder, 2003; Karreman & Steehouder, 2004; Kieras & Bovair, 1984; Patrick & Haines, 1988), but this definition can still refer to a broad category of information.

Kieras and Bovair (1984) found that principles were only effective if they were relevant to the tasks that the learners had to do. The principles they provided to their participants contained a few different types of information and not all were found to be helpful for procedural learning. Describing the device in the context of a familiar fantasy world (Star Trek) to increase motivation and providing general principles and design rationale did not seem to affect procedural learning. The information that was found to be helpful described how components were connected and how power was routed through the controls. The participants could use this information to infer procedures they needed to carry out as they understood how controls related to the internal components.

Later studies have taken advantage of the findings of Kieras and Bovair (1984) and most try to specify how the principles are relevant to the tasks used. Some researchers have used task analysis to define what information the principles should contain to make them relevant to the task (Catrambone, 1995; Kontogiannis & Sheperd, 1999). Often though, it is not clear how the principles are relevant to the tasks or whether the description of internal workings also has information on the relationship between controls and internal components (e.g., Borgman, 1999; Karreman & Steehouder, 2003; Karreman & Steehouder, 2004; Patrick & Haines, 1988).

It is not known what role the actual implementation of the principles might play in the discrepancy of findings in the research literature, but it is clear that to adequately test whether principles can help procedural learning it must be established that the information provided in the principles is relevant to the tasks the learners do. In addition, to apply the principles the learner must understand the information provided in the principle. That is, when using principles in procedural tasks learners must first understand the information and then figure out how to apply it. This means that if principles in instructions are to be helpful for procedural learning, the learners must be provided with the opportunity of understanding or learning the principle and applying it.

Procedural Instruction Specificity

Procedural instructions can be written at different levels of detail. Catrambone (1990) studied the effect of providing either general or specific procedural instructions to participants learning to use a word processing system. He defined specificity in a functional way by considering the number of cases the instructions covered. For example, the instructions to “log in” cover more cases than the instructions to “enter login name in the field marked ‘user name’, enter password in the next field and then click on the ‘log in’ button”. That is, more specific procedural instruction by definition become more restricted to specific cases of the task as they describe them in more detail.

Catrambone (1990) pointed out that more detailed instructions have the benefit of being unambiguous and make the task easier to complete initially as they describe exactly what needs to be done to accomplish the task. Detailed instructions would, however, be less effective for generalizing across tasks as learners have to abstract for themselves the common actions. Conversely, general instruction might be difficult to use initially, but be more helpful for generalizing the method across tasks. The results supported these hypotheses, and indicate that the specificity of procedural instructions is an important condition for determining learning and transfer.

Later, Catrambone (1995) showed that adding principles to general procedural instructions helped both initial performance and transfer of learning. By adding principles to general procedural instructions the learner gets best of both worlds; they can quickly start the task because the principles help them understand aspects of the general instructions that might otherwise be ambiguous, and transfer of learning is enhanced because the learners can reason effectively about how to carry out the new task.

Duff and Barnard (1990) demonstrated that there is a trade-off between procedural instructions and principles. In their experiment all the participants received procedural instructions and half also received principles. For the participants using principles the number of tasks where they received detailed procedural instructions was varied. The results showed that the beneficial effects of principles on procedural learning depended on how often they were supplied with detailed procedural instructions: The fewer the tasks with detailed procedural instructions, the more participants benefited from using the principles. Duff and Barnard concluded that learners benefit from principles only if they are forced to use them during practice, and only then will learners develop mental representations sufficiently robust to allow inferences when faced with new tasks.

Taken together these studies indicate that procedural instruction specificity might determine whether learners use the principles provided in instructions. Catrambone (1990, 1995) demonstrated that principles could help learners infer what to do when using more general instructions, and Duff and Barnard (1990) showed that principles seemed helpful only when the learners had less access to detailed procedural instructions. These conclusions invoke a specific assumption about how people use instructions, one that Duff and Barnard stated explicitly: people rely on the mental representation that most readily allows action execution. This means that if people faced with unfamiliar tasks are provided with enough information to determine action, they will make less use of inferential processing. Therefore, if procedural instructions are detailed users will have little need to consider any other information provided, but if the procedural instructions are general they will need more information to infer exactly what action to take.

Conclusion: Principles and Procedural Instruction Specificity

In prior research on principles in instructions, researchers often do not provide detailed descriptions of how the principles were derived and implemented. This can become problematic given prior findings on the importance of making sure that the principles are relevant to the task, especially by including in the principle information about how the controls relate to the inner workings of the system. Some researchers have tackled this problem by using task analysis when creating the instructional materials, thereby guaranteeing that the information in the principles is relevant to the tasks that the learners have to do.

Evidence suggests that the details provided in the procedural instructions might influence whether and how the principles are studied. Providing learners with general rather than detailed procedural instructions should encourage them to study and rely on the information in the principles while figuring out how to complete the tasks.

Research Question

The mental model account of how principles can help procedural learning assumes that providing principles leads to better knowledge or understanding of how the system works and this in turn will assist the learner in using the system. The research and theoretical approaches discussed above indicate that even if this account could be valid the situation is probably more complicated as there are factors that presumably influence

(1) whether principles lead to better knowledge or understanding of the system and (2) whether knowledge of the system assists the learner in using the system.

The first condition for determining whether principles can help procedural learning is to make sure that the principles are relevant to the tasks the learners need to perform. For example, a principle describing temperature for freezing water is not informative for learning how to ice-skate. Indeed, Kieras and Bovair (1984) demonstrated that the principles they provided their participants were helpful because they provided information that helped learners to complete the assigned tasks. Kieras and Bovair did not find beneficial effects of a motivational cover story or any other declarative information that was not directly relevant to the tasks the learners had to complete. In the same vein, research has indicated that theory is not useful unless it is demonstrated to the users how to put it to practice (Kontogiannis & Sheperd, 1999; Morris & Rouse, 1985; Patrick & Haines, 1988). This indicates that it matters whether and how the principles relate to the tasks that the users have to do. In the proposed research relevance of principles to tasks will be established by using task analysis to figure out what information learners need to carry out the tasks used in the experiments.

A second condition for determining whether principles can help procedural learning is to demonstrate the relationship between the knowledge of the system (or mental model) and procedural learning. The effect of the instructional method on mental models or knowledge must be measured separately from the effect of the instructional method on task performance. In the experiments proposed here the knowledge of the system (declarative learning) will be measured separately from procedural learning.

The aim of the current research agenda is to investigate in two experiments some factors that determine whether and how providing principles in user instructions is effective for declarative and procedural learning. Three factors have been identified as potentially influential when investigating whether principles are helpful for learning: The timing of principle use, the method used to study the principles, and the procedural instruction specificity. These factors were addressed in the two experiments.

The timing of principles use was the focus of the first experiment, specifically, to compare the effects of studying the principles before, during, or after completing the training tasks on declarative and procedural learning. This comparison was designed to determine whether it would be beneficial for learning (both declarative and procedural) to study the principles before engaging in the training tasks (as would be suggested by research on advance organizers), during training task completion (as would be suggested by JIT and the FLOABN model), or after task completion (which could be beneficial if learners make the effort to assimilate the information to the experience of doing the tasks).

The method of studying the principles was also addressed in the first experiment. As has been described, learning seems in large part dependent on how the to-be-learned information is processed cognitively, and that effortful cognitive processing (such as problem solving and assimilation) are important aspects of the learning processes. This means that passive methods such as reading should result in worse learning outcomes when compared to methods that require the learner to actively engage with the to-be-learned materials (Chi, 2009). The active study method used in the experiments was to ask the participants to summarize the main ideas in the principles. This method was chosen for two reasons: First, it is an active study method which is generally considered effective (Brown, Campione, & Day, 1981; Chi, 2009; Kirkland & Saunders, 1991; Pressley, 2006). Second, the summaries themselves provide a way to examine how the participants are studying the principles. The comparison between reading and summarizing the principles was expected to demonstrate differences in both declarative and procedural learning, as the participants who are required to engage in an activity (summarizing) were expected to learn the information provided in the principles better as they actively engaged in the material whereas the participants who simply read the principles were not expected to retain the information very well. This was expected to influence learning as the participants in these two groups only had access before starting the training tasks, but not while they did so; if the participants had not learned the principles they would not be able to apply them. The result of this comparison was expected to demonstrate the importance of either controlling or measuring how the instructional materials are used in studies of this kind.

What kind of procedural instructions are used was addressed in the second experiment. Specifically, the effect of using detailed or general procedural instructions was compared. Prior research has indicated that the specificity of procedural instructions influences how the principles are used, and in the second experiment it was investigated whether the learning situation can be structured (through specificity of procedural instructions and when principles are provided) to get learners to adopt a strategy of instructional usage that determines learning outcomes. The participants who used detailed procedural instructions were expected to report little cognitive load in training and show good training performance as compared to participants who used general procedural instructions, but the opposite pattern of results was expected for testing, and show that using detailed procedural instructions is detrimental for both declarative and procedural learning.

Overview of Experimental Methods

In both experiments, participants learned to troubleshoot a simulated electric circuit (Simutech, 2009). In the simulation, the tasks involve finding and repairing different faults in a circuit where pushbuttons and a relay control two light bulbs. The simulation provides a good platform for the experiments because it is a system with well-defined principles and because skill at troubleshooting electrical faults relies on understanding how the system works and how to find the faults. At the same time it is possible to provide clear procedural instructions to repair a fault.

Both experiments used the same set of seven troubleshooting tasks. In four of these the circuit was not operating properly because there was an "open" in the circuit (a break in the current path) and in three tasks the problem was due to a short in the circuit (the current travels on a different path than intended). Participants completed three training tasks with instructions and then all tasks without the instruction at testing. Tasks they completed during training were used to test retention and tasks they had not seen before were used to test near transfer.

All instructional materials for both experiments were created from a task analysis of the tasks used in the experiment. The task analysis was used to determine what information the learners needed to carry out the tasks and both procedural instructions and principles were based on this information. The detailed procedural instructions were created first, and then a predefined algorithm was used to abbreviate the detailed procedural instructions into a general version (see Appendix A for details on how the instructions were created).

EXPERIMENT 1

Introduction

The first experiment was designed to answer two questions: First, does the timing of principle use influence declarative and procedural learning? That is, does it matter whether learners study the principles before, during, or after completing training tasks? Second, does the method used to study the principles (summarizing or reading) influence declarative and procedural learning?

Research has indicated that if provided with general procedural instructions learners will have to figure out what the instructions refer to and how to carry out the directives, that is, some level of problem solving is needed. In the first experiment all participants received general procedural instructions when completing the training tasks to require them to engage in problem solving during training. The participants also received principles, but the timing of their use and method of study were varied, creating four groups.

The participants in the *Read-Before* group read the principles before starting the training tasks, but did not have access to the principles during training task completion. The participants in the *Summarize-Before* group studied the principles in the context of an activity by summarizing the main ideas in each principle before doing the training tasks, but also did not have access to the principles while completing the training tasks. The participants in the *Use-During* group only had access to the principles while completing the training tasks, but were not required to study them outside of training. The participants in the *Summarize-After* group did not have access to the principles until after completing the training tasks and then had to summarize the principles.

Comparing the Summarize-Before, Use-During, and Summarize-After groups provided insight into the effect of the timing of principle use, whereas a comparison of the effect of study methods was based on the Summarize-Before and Read-Before groups. I did not include a manipulation where participants would read the principles after (i.e., a Read-After group) because the Summarize-After group was included as a control comparison to show training performance without the principles. Also, the comparison of study methods only requires two groups that differ on study method and not timing of principle use. Using the study before manipulation was selected because instructions, and especially principles, are usually included with the assumption that learners study the information before engaging with the tasks. Therefore, the comparison between summarizing and reading principles before doing the training tasks was considered an appropriate comparison of the influence of study method.

All participants were tested on declarative knowledge of the system and procedural learning (performance on troubleshooting tasks) after completing the training and I expected different outcomes based on the experimental condition.

Participants in the Summarize-Before group, who summarized the principles before completing the training tasks, were expected to show good knowledge of the system because they would actively engage in learning the information when summarizing. These participants were also expected to be able to use the principles as advance organizers, to help them organize their knowledge and provide context for their problem solving experiences during training, and as a result their procedural learning would increase (at least in comparison to the Read-Before and Summarize-After groups). However, their procedural learning was expected to suffer (at least in comparison to the Use-During group), as their knowledge of the principles would not be organized in the context of doing the tasks and the process of trying to remember the relevant information from the principles and then apply it was expected to be effortful.

The participants in the Use-During group, who used the principles during training task completion, were expected to acquire good procedural learning because they studied the principles in the context of problem solving and the information in the principles was provided just-in-time, or when they needed it. The principles would therefore be encoded in the process of creating procedural knowledge and this was expected to lead to good procedural understanding. However, because the principles were studied in the context of the task, these participants were not expected to encode the information as organized declarative information. The participants in this group were therefore expected to demonstrate good procedural learning, but their declarative learning was expected to be worse than for the participants in the Summarize-Before and Summarize-After groups.

The participants in the Summarize-After group had only the general procedural instructions to rely on while completing the training tasks and their performance during training served as a baseline for how the tasks could be completed without the principles. After they completed the training tasks they studied the principles by summarizing. By studying the principles after completing the training tasks, the information in the principles would be encoded independently of doing the tasks, but after having had task experience. It was not clear which of these factors would be dominant; if the information would be encoded independently as declarative information and little effort made to integrate it with the problem solving experience (prior knowledge at this point in time) then good declarative learning but poor procedural learning would be expected. But if they would make an effort to integrate the information into the procedural knowledge gained during problem solving then better procedural learning could be expected. However, procedural learning was generally expected to be worse for the Summarize-After group than for the Summarize-Before group, because the latter group got an opportunity to apply the principles during training.

Participants in the Read-Before group, who read the principles before completing the training tasks, were not expected to encode or retain the information in the principles very well. They were also expected to have problems during training task completion because they would not have the information available in memory to help them during problem solving when using the general procedural instruction to complete the training tasks. I expected both declarative and procedural learning would be worse for the Read-Before group compared to the Summarize-Before group.

Based on these expectations I had five main hypotheses regarding the learning outcomes for these experimental manipulations (see Table 1 for an overview of the hypotheses).

Table 1
An overview of expected outcomes for each condition, reflecting the predictions of the six main hypotheses.

Condition	Training task performance	Declarative learning (knowledge of the system tests)	Procedural learning (testing task performance)	Subjective workload and task difficulty
Read-Before	Next worst	Worst	Worst	Worst (testing)
Summarize-Before	Next best	Best	Next best	Next best
Use-During	Best	Worst	Best	Best
Summarize-After	Worst	Next best	Worst	Worst (training)

Hypothesis 1 (timing of principle use and declarative learning): Participants who summarized the principles before or after completing the training tasks (Summarize-Before and Summarize-After) were expected to show better knowledge of the system than the participants who used the principles during training (Use-During). In addition, if the principles act as advance organizers then the participants in the Summarize-Before group would show better knowledge than the participants in the Summarize-After group.

Hypothesis 2 (the method of studying the principles and declarative learning): Participants who summarized the principles before doing the training tasks (Summarize-Before) were expected to show better knowledge of the system than the participants who read the principles before completing the training tasks (Read-Before).

Hypothesis 3 (timing of principle use and procedural learning): Participants who used the principles during training (Use-During) were expected to show better procedural learning than the participants who summarized the principles before completing the training tasks (Summarize-Before), which in turn were expected to show better procedural learning than the participants who summarized the principles after completing the training tasks (Summarize-After).

Hypothesis 4 (the method of studying the principles and procedural learning): Participants who summarized the principles before training (Summarize-Before) were expected to show better procedural learning compared to the participants who read the principles before training (Read-Before).

Hypotheses 1 and 3 were expected to demonstrate that the timing of principle use would influence declarative and procedural learning differently: Having good declarative knowledge would not mean knowing how to apply the information when completing tasks, but studying the principles in the context of applying them would lead to better procedural learning.

Hypotheses 2 and 4 were expected to show the importance of the method of studying the principles and knowing what kind of processing or learning activity learners actually employ. In previous studies in this domain, the way in which participants study the information has varied considerably, and the goal here was to demonstrate that the results of studies that require the learner to study the principles in a certain way or engage in a certain type of processing activity (Catrambone, 1995; Kieras & Bovair, 1984) are not necessarily comparable to studies that do not require the learners to engage with the materials in a particular way (Karreman & Steehouder, 2003; Karreman & Steehouder, 2004).

Hypothesis 5: Training performance was predicted to be, in the order of best to worst, Use-During, Summarize-Before, Read-Before, and Summarize-After. An important part of the experimental design was to make sure all the participants, regardless of experimental condition, would have had access to the same information before testing (both procedural instructions and principles). The Summarize-After group provided a baseline measure of how the training tasks could be completed with only the information from the general procedural instructions, and I therefore expected the participants in this group to show the worst outcome in training compared to the other groups. I believed comparing the training outcome of the Summarize-After and Read-Before groups would show some benefit of having at least read the principles before starting the training tasks. I expected the Read-Before group to have worse performance during training than the Summarize-Before group because the Read-Before participants would not have studied the principles as well and would therefore not be able draw upon that knowledge to help them during training task completion. The Summarize-Before group was expected to have worse training performance than the Use-During group because the Summarize-Before participants would have to rely on their memory of information learned from the principles to assist them during training whereas the participants in the Use-During group would have access to this information while solving the training tasks.

Hypothesis 6: Ratings of subjective workload and task difficulty were expected to be lowest for the Use-During group both during training and testing, but highest for the Summarize-After group during training, and Read-Before group during testing. The participants in the Use-During group were expected to report lower subjective workload and task difficulty during training because they had both principles and procedural instructions to assist them. In addition, I expected better procedural learning from them compared to the other three groups and that they would report less subjective workload and task difficulty during testing as a consequence. The participants in the Summarize-

After group were expected to report the highest workload and task difficulty during training because they had the least amount of information to assist them. However, in testing the participants in the Read-Before group were expected to report higher subjective workload and task difficulty than the other groups because they would have the worst memory of principles and therefore less information to assist them in solving the tasks.

Method

Participants

Ninety-nine undergraduate students at the Georgia Institute of Technology were recruited for the experiment. Students who had taken the course ECE 2040 at Georgia Tech, which covers circuit analysis, were not eligible to participate. The age range of the participants was between 18 and 24 years (M = 20, SD = 1.5) and 60 were male (60.6%).

The participants were randomly assigned to an experimental condition and compensated with course credit. This resulted in 26 participants being assigned to the Read-Before condition, 25 to the Summarize-After condition, and 24 to the Summarize-Before and Use-During conditions. During data collection it became apparent that the data from three participants (two in the Read-Before condition and one in the Summarize-After condition) could not be used because these participants had not been able to complete the first testing session and extra participants were recruited for these two conditions. Therefore, at the end of data collection there were 24 participants in each condition for a total of 96 participants in the experiment. Four participants could not return for the second session (two were in the Use-During condition, one was in the Summarize-Before condition, and one was in the Summarize-After condition).

Design

The experiment had a between-subjects design, where each participant experienced a single level of the experimental manipulations.

There were two independent variables in the experiment: timing of principle use and study method. The timing of principle use was manipulated by having the participants study the principles before, during, or after engaging in the training tasks. These were the only times that the participants had access to the principles. Study method refers to whether the participants were asked to study the principles by reading or summarizing them.

There were four different conditions in the experiment. The first group (*Summarize-Before*) had to actively study the principles before doing the training tasks by summarizing the main points of the principles. These participants did not have access to the principles while completing the training tasks (and knew that beforehand). The second group (*Use-During*) did not see the principles before the training, but had access to them while completing the training tasks and could therefore use the principles during problem solving. The third group (*Summarize-After*) did not have access to the principles until after finishing the training tasks, when they were asked to actively study the principles by summarizing the main points. The fourth group (*Read-Before*) was required to read the principles before doing the training tasks, but did not have access to the principles while completing the training tasks.

Because of the theoretical expectations the independent variables were not completely crossed, but instead considered in parallel. The comparison of the Summarize-Before, Use-During, and Summarize-After groups was intended to investigate the effect of timing of principle use, whereas the comparison of the Summarize-Before and Read-Before groups was intended to investigate the effect of study method.

Multiple dependent measures were used in the experiment. At the start of the experiment, participants were asked for demographic information (gender, age), GPA and SAT scores, and asked to complete the VARK learning style questionnaire (Fleming, 2009), assessing their preferred learning modality. During training the use of instructions was measured with the time spent studying each instructional page, both principles and procedural instructions. Procedural performance (both during training and testing) was measured with performance on troubleshooting tasks and involved four different dependent measures: time-on-task, number of safety errors (errors resulting in a virtual electric shock and restart of the task), number of unnecessary components replaced (the circuit was repaired by replacing the faulty component – replacing a working component therefore represented an unsuccessful attempt at completing the task), and number of meter readings (voltmeter and ohmmeter) used to complete the tasks. These measures were standardized and then the standardized scores were averaged to create a single standardized performance measure. After each task had been completed participants were asked to describe what had been wrong with the circuit (these answers were rated for accuracy) and asked to rate subjective workload (with abbreviated NASA-TLX) and task difficulty. In addition, whether participants needed help to complete the task (hints) or could do so on their own was recorded.

After training, declarative learning was measured with tests of knowledge of the system administered at two different occasions. At the end of the experiment, participants drew the circuit from memory and answered three questions about the workings of the circuit.

The experiment was in three sessions. In the first session (*Training Session*) the participants experienced the experimental manipulation and completed the training tasks with the aid of the instructions (where performance on training tasks and instructional use was measured). The second session (*Immediate Testing Session*) took place directly after the training session and required the participants to complete a test on system knowledge and three testing tasks (two new tasks and one that had been used in training) without the aid of instructions (performance on testing tasks and knowledge test was measured). The third session (*Delayed Testing Session*) was identical to the second session except that it took place a week after the training and included different testing tasks (two new tasks and two that had been used in training).

Materials

The experiment was computerized: The system, tasks, and instructions were all implemented on a computer. The measures completed for every task were computerized (NASA-TLX and ratings of task difficulty) but the preferred learning modality questionnaire (VARK) and the knowledge tests were on paper.

Simulation

The participants learned to troubleshoot a simulated electrical circuit using Simutech's Troubleshooting Electrical Circuits (Simutech, 2009). This simulation was created to teach electricians and others working with electrical circuitry troubleshooting skills in a safe environment. Participants interacted with the system using only the mouse control.

In the simulation (which was displayed on a 15" PC monitor) the participants saw a simulated electric circuit consisting of two light bulbs, a fuse, three ON-buttons (pressing any of them turned the lights on), three OFF-buttons (pressing any of them turned the lights off), and a relay (see Figure 1). Beside the circuit was a breaker panel where the current to the circuit could be turned off and locked out (the switch controlling the current to the circuit is sealed in the off position). In real life locking out the current is an important safety measure as it prevents someone accidentally turning the current on while the circuit is being worked on.

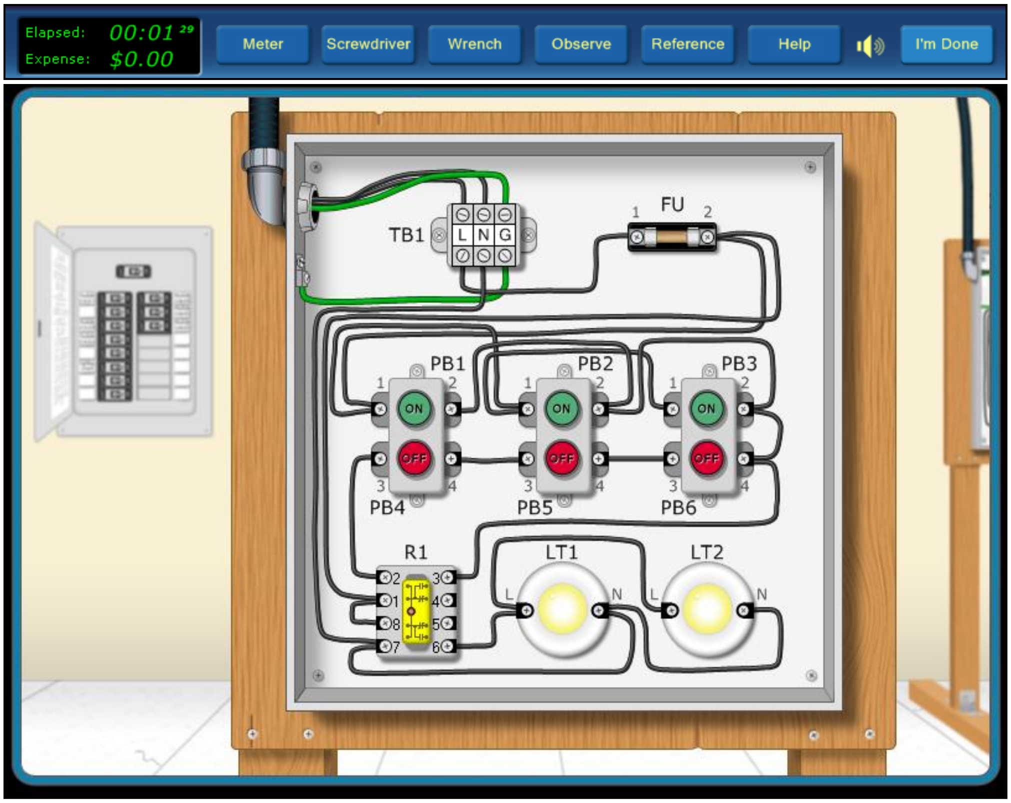


Figure 1. A screen shot of the electrical circuit simulation, with the lights turned on.

The top panel of the simulation showed the elapsed time and the accrued cost of repairs and provided access to various tools. A meter was provided to measure voltage and resistance, a screwdriver to loosen or tighten wires, and a wrench to replace components. In addition, a user could have access to an observing tool, diagram references, and a built in help system, but in the experiment the participant were not allowed to use those tools (the buttons were covered).

When the circuit operates normally, pressing any of the ON-buttons (PB1, PB2, and PB3) completes the circuit and energizes the relay (R1), then the two seal-in contacts in the relay (R1:1-3 and R1:8-6) close, and the two lights (LT1 and LT2) are energized. The seal-in contacts allow the relay to stay energized and the lights to stay on even when the ON-buttons are released again. When the OFF-buttons (PB4, PB5, and PB6) are pushed the relay becomes de-energized. By de-energizing the relay the two seal-in contacts disconnect (open) and de-energize the lights. When the OFF-button is released the lights stay de-energized because the seal-in contacts of the relay are now open (see Figure 2 for a wiring and schematic diagrams of the circuit).

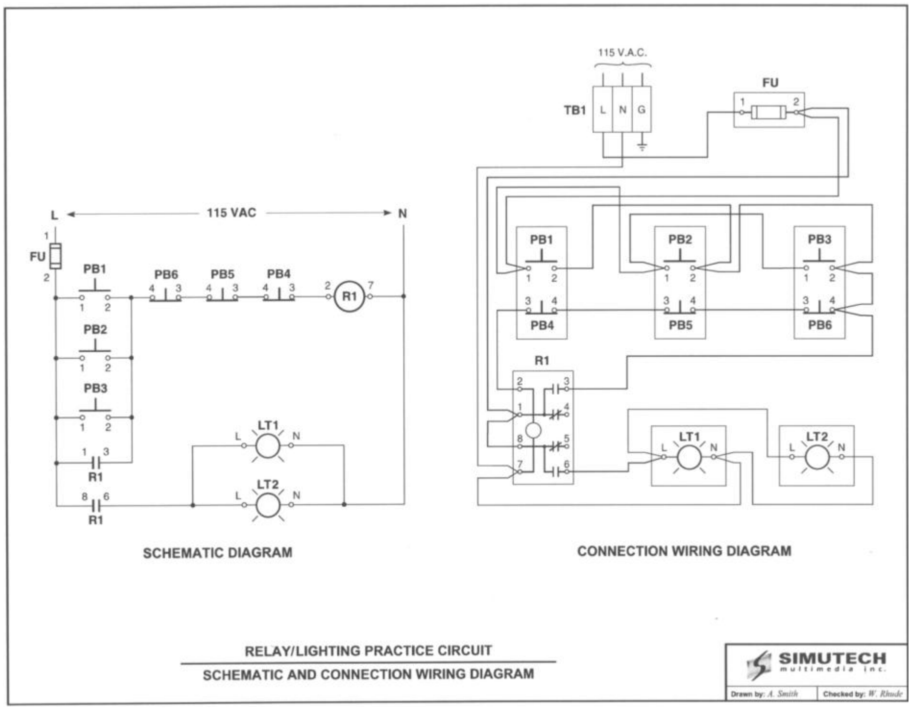


Figure 2. Schematic and wiring diagram for the circuit. The schematic diagram shows the functionality of the circuit, whereas the wiring diagram depicts the actual wiring used.

<u>Tasks</u>

The participant's task was to find and repair a fault in the circuit. The faults were either the result of a break in the circuitry (an open) preventing current from flowing, or a short to ground (a short) which occurs when two or more isolated components come into contact. In every case a particular component or a wire was faulty and needed to be replaced. The act of replacing the faulty component of wire therefore fixed the circuit.

All seven tasks used in the experiment were different, four were caused by an open somewhere in the circuit and three were caused by a short (see Table 2).

Table 2
Description of the tasks used in the experiment. The terms "allowed" refer to the amount of time and number of meter readings someone is allowed before he or she is considered unsuccessful at a task.

Task	Cause	Allowed task time	Allowed number of meter readings	Description
O1	Open	15 min	8	Wire loose on terminal PB5-3
O2	Open	15 min	4	Relay coil is open
O3	Open	15 min	6	Wire loose on terminal TB1-L
O4	Open	15 min	4	Relay contact R1-8 to R1-6 is open
S1	Short	20 min	7	Wire between PB2-1 and PB3-1 is shorted
S2	Short	20 min	8	Relay coil is shorted
S3	Short	20 min	10	Wire between PB6-4 and R1-3 is shorted

In the simulation, to be considered successfully completed tasks have to be finished within certain amount of time and with fewer than allowed number of meter readings (referred to as par-values). These par-values are based on how an expert electrician using the troubleshooting method advocated by the makers of the simulation solves these faults. The par-values therefore represent a reference point to what domain experts would consider acceptable performance (W. Rhude, September 18, 2009, personal communications).

In the experiment the par-values for maximum time needed to complete the task were used to define the time before participants had access to hints to help them solve the tasks (the hint structure and implementation are described later).

The par-values for meter readings were used to define the minimum number of meter readings participants would be expected to need, to exclude cases where few or no meter readings had been logged. The simulation only counted the correct placement of the electrodes as meter readings, and as a consequence, some participants who did not use the meter correctly had very few and even zero meter readings logged for some tasks. Therefore, I filtered out cases with less than four meter readings logged on the assumption that participants in the experiment (domain novices) should be expected to need more meter readings than experts and the lowest number an expert would need for any of the tasks was four meter readings. A total of 81 task instances were removed (9% of the total), 21 from Read-Before, 17 from Summarize-Before, 18 from Use-During, and 25 from Summarize-After.

Each participant completed three training tasks and the four remaining tasks were used for testing. There were two sets of tasks that were counterbalanced in each condition. The first set (set A) had O1, O4, and S3 as training tasks, and the second set (set B) had O3, O2, and S1 as training tasks. The two sets of tasks were complementary in terms of what they teach and test (the type of fault and method required for completion), but not identical (see Appendix B for a detailed description). This means that the tests of transfer would be considered tests of near transfer as the transfer is within the same domain and the same context (Barnett & Ceci, 2002). The seventh task, S2, was only used for testing transfer, because it involved a slightly different fault than encountered in other tasks (a short in the relay).

As well as being tested on new tasks (transfer tasks) the participants were tested on the tasks they completed in training (retention tasks). They were tested on one of the training tasks in the immediate testing session, and the other two in the delayed testing session. Therefore, the participants completed three testing tasks in the immediate testing session and four testing tasks in the delayed testing session.

There were four different task orders for each set of training tasks covering every combination of testing tasks (with some constraints) creating a total of eight task orders in the experiment (see Appendix B for further details). This was done to test for effects of task order.

Instructional System

Participants had access to general procedural instructions and principles to help them during the training session, but they did not have access to any instructions during the two testing sessions.

The instructions were presented on a separate computer from the simulation, and the participants could see both the simulation and the instructional system at the same time.

When doing a task the participants saw the main page of the instructions which displayed the titles of the different instructional pieces the participants could view. To read each section the participants had to click on the button for that piece of information and hold down the left mouse button to view it. As soon as the mouse button was released the main page of the instruction system appeared again (see Figure 3 for examples of screenshots).

This implementation of the instruction system made it impossible for the participant to work on the simulation and view the content of the instructions at the same time, and allowed me to record the time spent studying the instructions.

On the main page of the instructions the titles of the different instruction sections indicated whether they contained procedural instructions or principles. Procedural instructions had titles indicating tasks (e.g., “Training task #1”), whereas principles had titles indicating which part of the system they referred to (e.g., “Voltmeter” and “Fuse”).

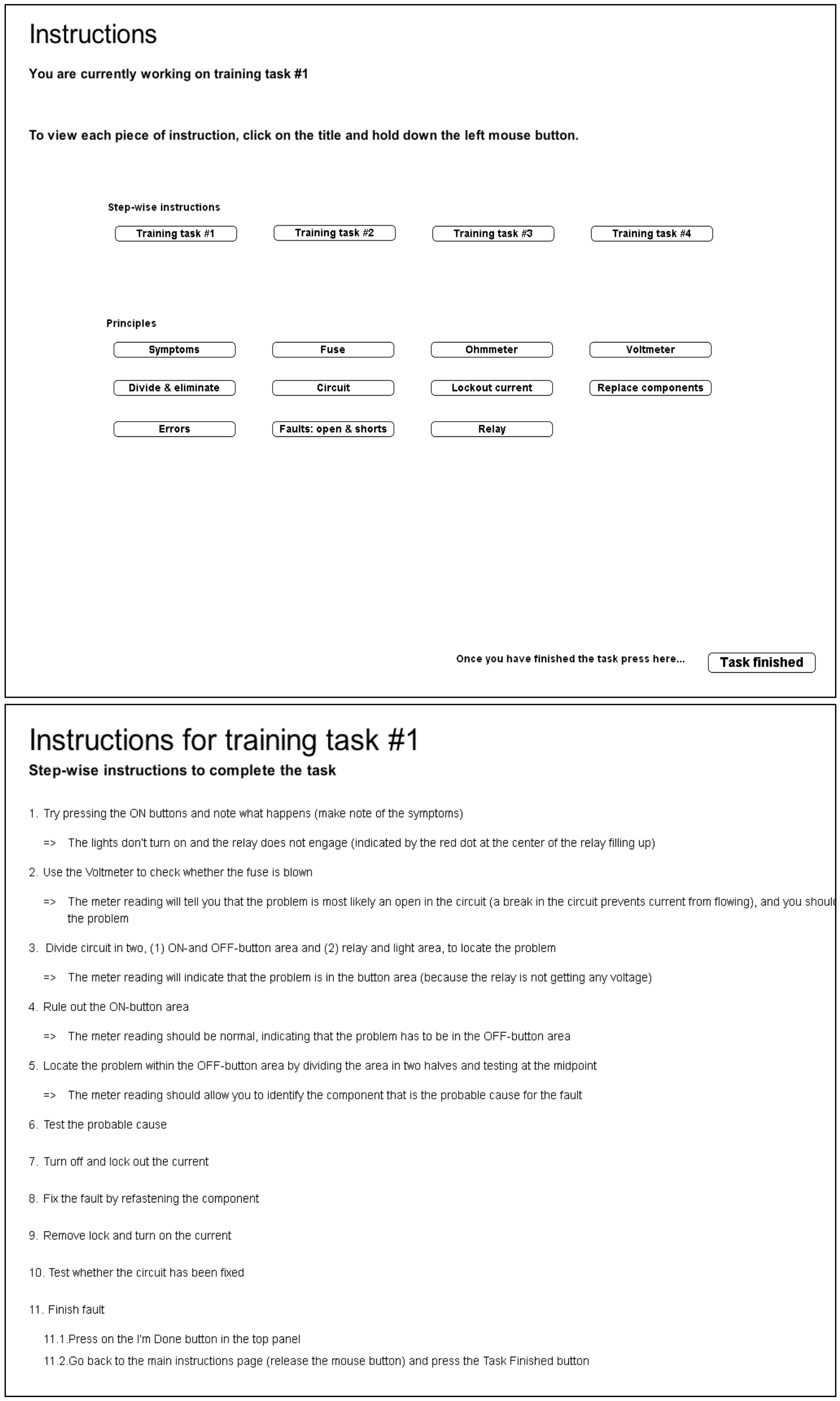


Figure 3. Screenshots of the instructional system. The panel on the top is the main page (as seen by a participant that has access to both procedural instructions and principles), and the lower panel shows an example of procedural instructions being viewed when the left mouse button is being held down.

Instructions

In the first experiment participants received principles and general procedural instructions. Both sets of instructions were created from a hierarchical task analysis to identify the information and actions needed to complete the tasks used in the experiment (the method of creating the instructions and examples of the instructions are provided in Appendix A). The principles described generalities and rules in the domain (on how to troubleshoot an electrical circuit) relevant to the tasks used in the experiment. The general procedural instructions provided general descriptions of each step needed to complete a task. The descriptions were general in that they did not include any sub-steps and interim conclusions, but only described the main part of the step and the final conclusion. In addition, earlier steps included more information than later steps, requiring the participants to engage in more problem solving as the task progressed. This also prevented the participants from reading only the last step of the procedure to know which component was faulty and needed to be replaced.

Hint System

The hint system was implemented after pilot testing showed that some participants had difficulty completing the tasks, especially the testing tasks where they did not have any instructions to aid them. The goal of the hint system was to make sure all participants completed each task for both practical (so the simulation would log all the task information) and methodological reasons (to guarantee equal learning opportunities for each task; see Appendix C for detailed description of the hint system and its development). In addition, the hint system allowed me to objectively define when participants failed to complete a task; if they required a hint to tell them exactly what was wrong with the circuit to fix it then they could be considered to have failed at completing the task on their own.

The hint structure for the testing tasks included three hints for each task of increasing specificity and they appeared one at a time after a certain amount of time had elapsed. The first hint told the participant whether the problem was an open or a short and which meter to use, the second hint appeared three minutes after the first and indicated the general location of the problem, the third hint appeared three minutes after the second one and described what the fault was and which component had to be replaced to fix it. Each hint was associated with a button that appeared on the main instruction page and the participants had to press the button to view the hint.

The hint system was developed for both training and testing tasks and was identical in both cases except for the time elapsed until the first system was shown. For the testing tasks the first hint did not appear until after the defined maximum par-value time (see Table 2) had elapsed for each task. For example, for a testing task with 15 minute maximum par-value the hints would not start appearing until after 15 minutes had elapsed. For the training tasks the first hint did not appear until after 40 (open) or 45 (short) minutes had elapsed.

Knowledge Tests

Knowledge of the system was measured with both multiple-choice questions and questions requiring a short written answer (open-ended questions). The knowledge questions assess knowledge of information provided in the principles and how the circuit works (see Table 3 for examples of knowledge questions).

Two equivalent but different versions (versions X and Y) of the knowledge questionnaire were created, one for each testing session. The questions for the two versions were written in pairs, such that each pair involved the same knowledge piece from the principles. The two versions were counterbalanced: half the participants in the Read-Before condition received version X in the first session and the other half received version Y. In the second session those who received version X in the first session got version Y and vice versa.

Alternate form reliability was assessed by calculating the Pearson *r* correlation between the two versions of the knowledge test for all the participants. As the two test versions were counterbalanced within each condition a correlation between the two versions should provide a measure of alternate form reliability without testing occasion becoming a confound. The correlation between the two versions of the knowledge test was high ($r = .801$; $p < .001$), indicating good alternate form reliability between the two versions as the criterion adopted was a correlation coefficient of .8 or higher (Nunnally, 1967).

Table 3
Examples of the knowledge questions.

Questions
If the fault can be anywhere in the circuit, how would you start dividing the circuit for testing according to the divide and eliminate approach? a. Into (1) supply and buttons, and (2) relay and lights b. Into (1) supply and fuse, and (2) ON-and OFF-buttons c. Into (1) relay, and (2) lights d. Into (1) ON-buttons, and (2) OFF-buttons
What is the purpose of the seal-in contacts in the relay? a. They make sure that the lights stay de-energized when the OFF-buttons are released. b. They make sure that the lights stay de-energized when the ON-buttons are released c. They make sure that the lights stay energized when the ON-buttons are released d. Both a and c
What happens if you press one of the OFF-buttons (when the lights are on and the circuit is behaving normally)? Please be as specific as you can.
What is an open?
Where does the relay coil receive voltage (draw a picture of the relay and circle the terminal)?

The questions were pilot tested to determine difficulty (whether there were ceiling or floor effects). The scores of the three pilot participants had an average of 21.83 (*SD* = 3.22) for version X and 20.17 (*SD* = 5.06) for version Y. There was therefore no evidence of a floor effect (the lowest possible score is 0, and pure guessing on the multiple-choice questions results on average in a score of 5) or a ceiling effect (the highest possible score was 36).

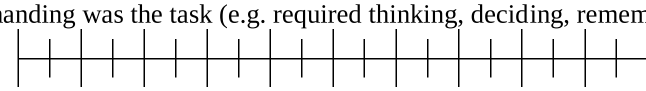
In addition, the mean differences of the scores for the two versions was compared within each testing session. When administered in the immediate testing session the mean scores of the two tests were very similar (X: $M = 22.93$, $SD = 4.55$; Y: $M = 21.88$, $SD = 5.25$) and there was no significant difference found between the two versions ($p > .05$). When administered in the delayed testing session the outcomes of the two tests were again very similar (X: $M = 22.27$, $SD = 4.89$; Y: $M = 22.27$, $SD = 5.86$) and there was no significant difference found between the two versions ($p > .05$). This suggests that the two forms provide comparable mean scores and are essentially identical in difficulty level.

Subjective Workload Questionnaire

Subjective workload was measured with an abbreviated NASA-TLX questionnaire (Langan-Fox, et al., 2006). The NASA-TLX is a self-report measure where participants rate their subjective workload on six different dimensions: mental demand, physical demand, temporal demand, success at task, difficulty of obtaining that level of success, and degree of frustration. Each of these dimensions is rated on the scale from 0 to 100 (see examples in Table 4).

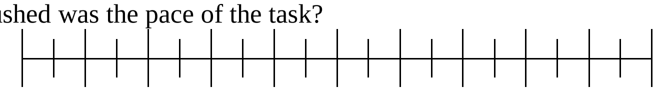
Table 4
Examples of items on the NASA-TLX subjective workload questionnaire.

1. How mentally demanding was the task (e.g. required thinking, deciding, remembering, searching)?



Very Low　　　　Very High

3. How hurried or rushed was the pace of the task?



Very Low　　　　Very High

Learning Styles Questionnaire

To assess the preferred learning modality of the participants the Visual-Aural-Read/write-Kinesthetic (VARK) questionnaire was used (Fleming, 2009). The VARK assesses the preference of the learner for taking in and disseminating information in a learning situation and categorizes the preferences as visual, aural, read/write, and kinesthetic. Participants answer 16 questions by selecting one or more statements that they feel applies best to them. From these selections the scores for each of the four categories are calculated. See Table 5 for item examples.

Table 5
Examples of items on the VARK learning styles questionnaire.

1. You are planning a holiday for a group. You want some feedback from them about the plan. You would:
    a. describe some of the highlights.
    b. use a map or website to show them the places.
    c. give them a copy of the printed itinerary.
    d. phone, text or email them

2. You are about to purchase a digital camera or mobile phone. Other than price, what would most influence your decision?
    a. Trying or testing it.
    b. Reading the details about its features.
    c. It is a modern design and looks good.
    d. The salesperson telling me about its features.

3. I like websites that have:
    a. things I can click on, shift or try.
    b. interesting design and visual features.
    c. interesting written descriptions, lists and explanations.
    d. audio channels where I can hear music, radio programs or interviews.

There is little information to be had on the validity of the VARK scale, but recent research by Leite and Svinicki (as reported in Fleming, 2009) has established adequate reliability for scores on the VARK subscales (0.85 for visual, 0.82 for aural, 0.84 for read/write, and 0.77 kinesthetic).

The main goal of using the VARK is to get a sense of the preference the participants have for receiving the information in a written format (read/write subscale) as compared to learning by doing or experiencing (kinesthetic subscale), and see whether this stated preference predicts learning outcomes. However, in light of recent review on empirical findings in the learning styles domain (Pashler, McDaniel, Rohrer, & Bjork, 2009), I do not predict that expressed preference for receiving information in a certain way will affect learning outcome. The review showed little evidence that a particular method of instruction is more effective given a certain learning style. The reason for including a measure of learning style preferences in this study is twofold: 1) to exclude possible alternative explanations and 2) see if stated preferences for learning modality predicted learning outcome as the experiment provided two very distinct modalities – learning from verbal written materials and learning from completing tasks.

Procedure

The experiment consisted of three sessions: A training session where participants completed three training tasks with the help of the instructions, an immediate testing session that started right after training was completed, and a delayed testing session that took place a week later. The procedure for the two testing sessions was identical, except at the end of the delayed session the participants were asked to draw the circuit and answer questions about the circuit (see Figure 4 for an overview of the procedure for the three sessions).

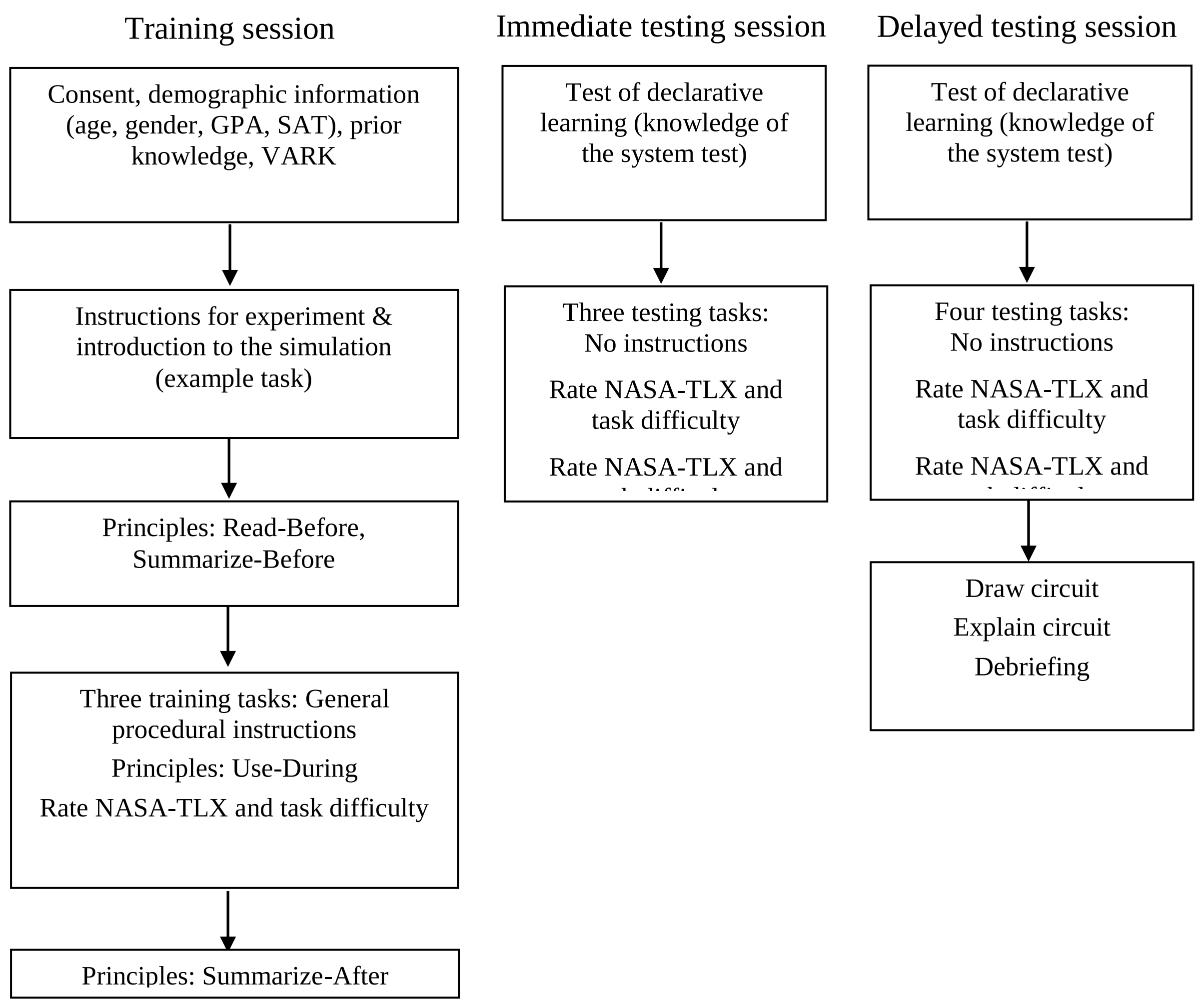


Figure 4. An overview of the procedure of the three sessions in the experiment. The immediate testing session took place immediately after the training finished whereas the delayed testing session took place a week later. The arrows represent the direction of time, and the events described at the top of the figure take place before the events described towards the bottom of the figure.

Overall, the participants had five hours to complete the whole experiment; three hours for the training and the immediate testing sessions, and two hours for the delayed testing session. Usually, participants needed about 90-120 minutes to complete the training, and 60-90 minutes to complete the first testing session. The second testing session took longer than the first because the participants had to complete one more testing task and the drawing and explaining assignments.

In the training session the experimental manipulation was implemented and the participants completed the training tasks with the aid of the instructions according to their assigned condition. In each testing session the participants completed knowledge tests, tasks they encountered during training (measuring retention) and unfamiliar tasks (measuring transfer). Every participant worked individually on two computers: One computer had the simulation and the other the instructional system.

Training Session

When starting the experiment, the participants were first asked to give informed consent. Then they were asked to provide demographic information (gender and age), their GPA and SAT scores, answer questions about prior domain knowledge, and complete the learning style questionnaire (VARK). Afterwards they were given instructions for the experiment itself (explaining what they had to do) and an introduction to the simulation.

In the instructions for the experiment the participants were provided with a description of the different sessions (training and testing) and told that they would only have access to instructions during training, but not testing. The goal was to motivate the participants to make an effort to learn during the training session.

The introduction to the simulation provided an overview of the simulated circuit and what the tasks entailed (e.g., finding a problem in an electrical circuit and fixing it). As a part of the introduction, the participants completed a simple example task (where one of the two light bulbs was broken) using detailed procedural instructions. They then rated the example task on subjective workload and difficulty as they later would the other tasks in the experiment. This was done both to familiarize the participants with the simulation and to provide an anchor for the rating scales. The participants were told that the example task would be the easiest task in the experiment and to rate it with that in mind. The plan was to use the rating of the example task as a baseline to compare against the ratings of the other tasks.

Once the participants had completed the example task the participants were asked to either study the principles (Summarize-Before and Read-Before) or start the training tasks (Use-During and Summarize-After) depending on the experimental condition.

The participants in the Read-Before group controlled how much time they used to study the principles, but the Summarize-Before participants had up to 45 min in total to read and summarize the principles. This upper limit was implemented as piloting indicated participants would otherwise spend over an hour summarizing the principles and extend the already long experiment. The participant had a specified amount of time to summarize each principle and this duration was displayed while the participants were summarizing. The upper limit was based on the pilot data and depended on the amount of information in each principle (see Appendix D for a detailed description of how the upper limit was calculated and implemented).

After having studied the principles the participants in the Summarize-Before and Read-Before groups started the training tasks. All participants completed three training tasks, one at a time, and had full access to the general procedural instructions while doing so. The participants in the Use-During group also had access to the principles while completing the training tasks.

For each training task the participants notified the experimenter that they were ready for the next task and the experimenter got the simulation ready. During task completion, the participants had access to the instructions on the instructional system, but could only view the details of the instructions while holding down the left mouse button. This allowed me to measure the time used to study the instructions.

To persuade participants to engage in troubleshooting and not, for example, replace every component to solve the task, they were asked to keep the "cost" as low as possible. The simulation shows accrued cost for each task, and the cost increases every time a component is replaced or a particular action performed (e.g., locking out the circuit or replacing a fuse). The cost of replacing a component is especially prominent because it was displayed where participants had to verify that they indeed wanted to replace the component. The cost therefore provided instant feedback to the participants on the efficiency of their troubleshooting and by emphasizing keeping the cost at minimum provided them with a clear definition of what constituted good performance.

After finishing each training task the participants were asked to complete the NASA-TLX questionnaire and rate the task difficulty of the task. In addition, the participants were asked to describe what was wrong with the circuit and what they did to fix it. This was done to gauge the understanding the participants had of the task they just completed.

After the three training tasks had been completed the participants in the Summarize-Before, Use-During, and Read-Before groups were finished with the training session, but the participants in the Summarize-After group summarized the principles before finishing the training session.

Testing Sessions

Each testing session started with the participants completing a knowledge test. After having completed the knowledge test the participants completed the testing tasks where they repaired the circuit without the aid of instructions. The testing tasks included tasks that the participants had completed in the training session (measure of retention) and new tasks (measure of near transfer).

There were three testing tasks in the immediate session; the first one was a task they completed in training (an open) and the other two were new (an open and a short). In the delayed session the participants completed four testing tasks; the first two were tasks they completed in training (an open and a short) and the other two were new tasks (an open and a short).

After completing each testing task the participants rated the subjective workload for the task (with the NASA-TLX questionnaire), rated the difficulty of the task, and described what had been wrong with the circuit. In addition, for the retention tasks the participants indicated whether they relied on memory to help them complete the tasks. All participants were notified which tasks they had already completed in training to provide them with the opportunity of relying on memory. I expected participants to prefer working from memory if possible because it would be a faster way to complete the task than relying on problem solving.

When the participants had completed all three testing tasks they were finished with the immediate testing session. However, when the participants had completed all the testing tasks in the delayed testing session they were asked draw a picture of the circuit from memory and then explain how the circuit worked by answering three questions. In the first question they were provided with a picture of the circuit and asked to write down the function and purpose of each component in the circuit. In the second question they were shown a close-up picture of the relay and asked to explain how the relay worked in detail. In the third question they were given a picture of the circuit and asked to highlight the sections that are energized when the lights are off and the circuit is functioning normally. These questions were meant to gauge the understanding or mental model of the circuit. For example, knowing where to expect voltage when the lights are off is essential in understanding how the circuit operates.

Preparing Data Analysis

Data Removal and Outliers

Out of the 96 participants (24 in each condition) the data from three could not be used. These three participants were all non-compliant as they pretended to have finished tasks when they had not. They could do this by dismissing the feedback window that appears after each task in the simulation. This feedback window specifies whether the task has been successfully completed. The participants were specifically told to leave this window on the screen after each task for the experimenter. If they did not, they were asked whether they had finished the task (i.e., fixed the circuit) and then asked again to leave the feedback window for the experimenter after the next task. Most participants complied except for these three participants, but it did not become apparent that the tasks had not been finished until after data collection had been completed. All of their data were removed, leaving 24 participants in the Read-Before condition and 23 in the other three conditions. Therefore data from 93 participants was used in the analyses for the first experiment.

Of the 93 participants, four could not return for the second session. Two were in the Use-During condition and one in the Summarize-Before and Summarize-After conditions respectively. This means that for the second session, there are in effect 21 participants in the Use-During condition, 22 in the Summarize-Before condition, 22 in the Summarize-After condition, and 24 in the Read-Before condition.

The difference in sample size was found to be small enough to not affect the precision of the estimate of the population mean because the difference in precision between the equal (intended) sample size (24 per condition) and the unequal sample size was 14% and anything less than 20% has been shown to have minimal effects on the precision on the estimate (van Belle, 2008). However, unequal sample size is always problematic (no matter how small the difference) if the variance among the groups compared differs, and Levene's test of equal variance was used to guarantee that the difference in sample size was not problematic. If the Levene's test is significant, the variance among the compared groups is not equal and parametric tests such as ANOVA and t-test cannot be used (the homogeneity of variance assumption is violated) (Keppel & Wickens, 2004).

Only one dependent measure had to be checked for outliers: task duration or time-on-task. Unusually long task times could indicate abnormalities in procedure, for example, the simulation computer needed restarting or the participant went to the bathroom. Abnormalities in procedure were documented in the participant log, but as this was not automated and as the experimenter was sometimes busy with other participants it cannot be guaranteed that all cases were entered into the log. Also, an extremely low value (under 10 seconds) almost certainly indicated an error (i.e., the “task finished” button pressed too early) as the tasks needed a minimum of 10-30 seconds to complete (an expert with prior knowledge of the faults in each task was timed completing the tasks). Task duration outliers were defined as being three standard deviations from the mean for the upper value and 10 seconds for the lower value. A total of 16 outliers (1.8%) were removed.

Coding

Coding was required for a few of the measures used in the experiment: the summaries created by the participants in the Summarize-Before and Summarize-After conditions, the open ended knowledge test questions, and the drawings and explanations of the circuit.

Summaries Created by the Participants

The summaries created by the participants in the Summarize-Before and Summarize-After conditions were coded based on whether they included the main ideas of each principle and whether they had been summarized in the participants own words or typed verbatim. The rationale behind the coding scheme was that a summary in the participant’s own words including all the main ideas conveyed in the principles would indicate better learning than an incomplete summary typed verbatim. In the first case the participant would be required to work with the learning materials and get engaged in processing the information at a deeper level than they would in the second case (Brown, et al., 1981; Chi, 2009; Pressley, 2006). Details about the coding of the summaries and interrater reliability are provided in Appendix E, along with the coding rubric.

Open Ended Questions on the Knowledge Tests

On the two knowledge tests, half the questions (18 of the 36 questions) were open ended and needed to be coded for accuracy. Each question was worth one point, but half a point could be given for partial credit. The details on the coding process, interrater reliability, and the coding rubric are provided in Appendix E.

Drawings of the Circuit

After completing the testing tasks in the second session the participants were asked to draw the circuit from memory in as much detail as they could. The main focus of the coding rubric was whether the participant had depicted all the elements (components and wires) correctly, then on details important for function (in terms of locating faults and finishing tasks), and lastly on details not relevant for functionality. Points were given for each of these components separately: showing the structure of the circuit (structure points), depicting functional details (functional detail points), and including general details in the drawing (drawing detail points). The details on the coding of the drawings and the coding rubric are provided in Appendix E.

Explanations of the Circuit

At the end of the delayed testing session the participants were asked to explain the workings of the circuit with three questions. The first question contained a picture of the circuit and participants were asked to provide a short explanation of the function of each component. The coding rubric for this question was based on the principles, and participants got points for how well they described the functioning of a component. In addition, for each component more points were given for advanced knowledge (e.g., the ON-buttons are by default open) than basic knowledge (e.g., the ON-buttons energize the relay and turn the lights on).

The second question showed a close up of the relay and the participants were asked to use the picture as a reference and explain in detail how the relay worked (specifically describing the relay coil and contacts). This question was added because pilot testing indicated participants found the relay complex and had difficulty understanding how it worked. Therefore, the understanding of this component could separate those with good knowledge of the circuit from those with less knowledge. This question was coded using the principle about the relay and how the circuit works as guidelines.

The third question showed a picture of the circuit and the participants were asked to use a red pen to mark where the circuit would be energized when the lights are off (assuming the circuit is functioning properly). They were given points for correctly highlighted parts, but points were deducted for incorrectly marked parts (zero points was the lowest possible score). The deduction was used so coloring every single wire and component in the circuit would not get full points; instead the full points were given for marking only the correct parts, nothing more and nothing less. The coding rubrics for the three questions and details of the coding process are provided in Appendix E.

Prior Knowledge

Prior domain knowledge was assessed before training in three ways: through self-reported prior experience, self-reported domain knowledge, and with a domain knowledge test. For the self-report of experience the participants were asked whether they had ever taken classes covering electrical circuits and whether they had ever worked on electrical circuits. In both cases participants were asked to describe their prior experience if they had any. For self-reported domain knowledge participants were asked to indicate whether they had no knowledge, some knowledge, or good knowledge of electrical circuits. The test to assess domain knowledge consisted of five questions, such as “what is an open in a circuit?”, “what does an ohmmeter measure?”, and “what does it mean when it is said that a component in a circuit is normally closed?”

Participant self-reported experience was coded as “no experience” (had not taken class or done any work), “some experience” (had either taken a class or done some work), and “experience” (had taken one or more classes and done some work). The majority of participants had no experience (57%), about third had some experience (30%), and a minority (13%) had experience. When self-reported experience of participants was compared for the four conditions, no difference was found among the groups ($p > .05$).

When rating their own knowledge of electrical circuits, 54% of participants said they had no knowledge of electrical circuits, 46% said they had some knowledge, but none of the participants said that they had good knowledge. There was no difference among the conditions in terms of self-reported domain knowledge ($p > .05$).

The scores on the five prior knowledge questions were coded either wrong (0) or right (1), by a single rater, and then aggregated into a score indicating domain knowledge. Half of all the participants scored low (0-1 on the test), 39% scored in the middle range (2-3), and 11% scored high (4-5). When the participants in the four conditions were compared on the prior knowledge score there was no difference among the groups ($p > .05$).

As the effects of prior knowledge on dependent measures would be analyzed using a regression, the level of multicollinearity among the three prior knowledge measures was tested by regressing all the variables on each other and checking the R^2 (a high correlation among the predictor variables can skew predictor calculations). A high R^2 (.8 or more) would indicate an unacceptable level of multicollinearity (Lewis-Beck, 1980). The categorical variables were dummy coded (e.g., no experience and no knowledge were coded with zeroes for comparison). None of the R^2 values were found to be problematic as all were below .8 (.42, .49, .44, .60) and the three measures could be used in a multiple regression to assess the effects of prior knowledge on the dependent variables.

Order Effects

Knowledge Test Order

Participants were presented with a knowledge test at each testing occasion, one immediately after training and one a week later. There were two versions of the knowledge test (version X and Y) and they were counterbalanced such that half of the participants received version X on the immediate session and half did so on the delayed session.

A t-test was used to determine whether the order of the knowledge test had an effect on the outcome on the test: whether it mattered that version X had been given in the first session and version Y in the second session (X-first) or vice versa (Y-first). No order effects were found (X-first: *M* = 22.87, *SD* = 4.60; Y-first: *M* = 22.35, *SD* = 4.80; $p > .05$ in both cases) and the order of the knowledge test was therefore ignored in the analysis of the results.

<u>Task Order</u>

The effects of the task set used for training was assessed by comparing the participants who received task set A in training with those who received task set B (half of participants in each condition). The order effects of the testing tasks were assessed by comparing the participants assigned to each order. The order effects were calculated for both knowledge of the system and task performance.

A comparison between participants being trained on task set A and task set B showed no difference in knowledge of the system (measured with the knowledge tests) or procedural learning (measured with performance on testing tasks); $p > .05$ in all cases. Therefore, the task sets were deemed equivalent and task set was not taken into account in further analyses.

Each participant was randomly assigned to one of eight predetermined order of the testing tasks (four for each task set). The order of the testing tasks did not have an effect on knowledge of the system or procedural learning ($p > .05$ in both cases) and therefore was not taken into account in the analysis of the results.

Results

The analyses addressed the six main hypotheses regarding the comparison on knowledge of the system (Hypothesis 1 and 2), procedural learning (Hypothesis 3 and 4), training performance (Hypothesis 5), and subjective workload and ratings of task difficulty (Hypothesis 6). Auxiliary analyses fell into four different categories: 1) Analysis of instruction use (i.e., how much the participants consulted the instructions), 2) comparing learning outcomes in the immediate and delayed testing sessions, 3) analysis of the quality of summaries made by the participants in the two summarizing conditions, 4) influence of preferred learning modality and demographic variables, and 5) analysis of drawings and explanations of the circuit.

After every task (both in training and testing) participants were asked to rate the difficulty of the task and recount what had been wrong with the circuit. In analysis these measures yielded very little information: The majority of participants knew what had been wrong with the circuit and rated the task difficulty as very high, and there were no differences among the conditions ($p > .05$ in all cases). Therefore to shorten the results, I moved all the analyses pertaining to knowledge of fault into Appendix F and all analyses on task difficulty into Appendix G.

In addition, participants were asked after completing each retention tasks (these were testing tasks they had already completed earlier in training), whether they had relied on memory of completing the task in training. An analysis showed no difference among the conditions ($p > .05$), but did reveal that the participants who relied on memory generally performed better on the retention tasks. These results do not speak directly to the hypotheses tested in the experiment and therefore I moved them to Appendix H.

Knowledge of the System (Hypotheses 1 & 2)

The knowledge of the system was measured on two occasions with two tests. Each test had 18 multiple-choice questions and 18 open ended questions, for a total of 36 questions. Each question was scored as either correct (1 point) or incorrect (0 points) and then the number of correct answers was summed, so that on each test a participant could score from 0 to 36 points.

As will be discussed in more detail later, no difference was found for knowledge test outcome on the two testing occasions (immediate session: $M = 22.63$, $SD = 4.65$; delayed session: $M = 22.57$, $SD = 4.78$; $p > .05$). Therefore, the results from the two testing occasions were collapsed for the analysis on knowledge test outcome. This means that every participant is associated with two knowledge test scores in the following analyses.

Levene's test of homogeneity of variance was calculated to guarantee an analysis of variance would be the appropriate statistic to use with the independent variable having unequal sample sizes. The test was not significant ($p > .05$) and the assumption of equal variance holds.

Timing of Principle Use (Hypothesis 1)

The first hypothesis predicted that participants in the Summarize-Before and Summarize-After groups would have better knowledge of the system as compared to the participants in the Use-During group because in these cases the participants would have a designated study period and method. In addition, if the principles act as advance organizers, the participants in the Summarize-Before group would be expected to show better knowledge of the system than the participants in the Summarize-After group.

Figure 5 shows the mean score for all four conditions (note that the figure includes the Read-Before group for completeness even if this group is not included in the current hypothesis). It is interesting to note that the difference between the highest scoring group (Summarize-After) and the lowest scoring group (Use-During) was only four questions on average.

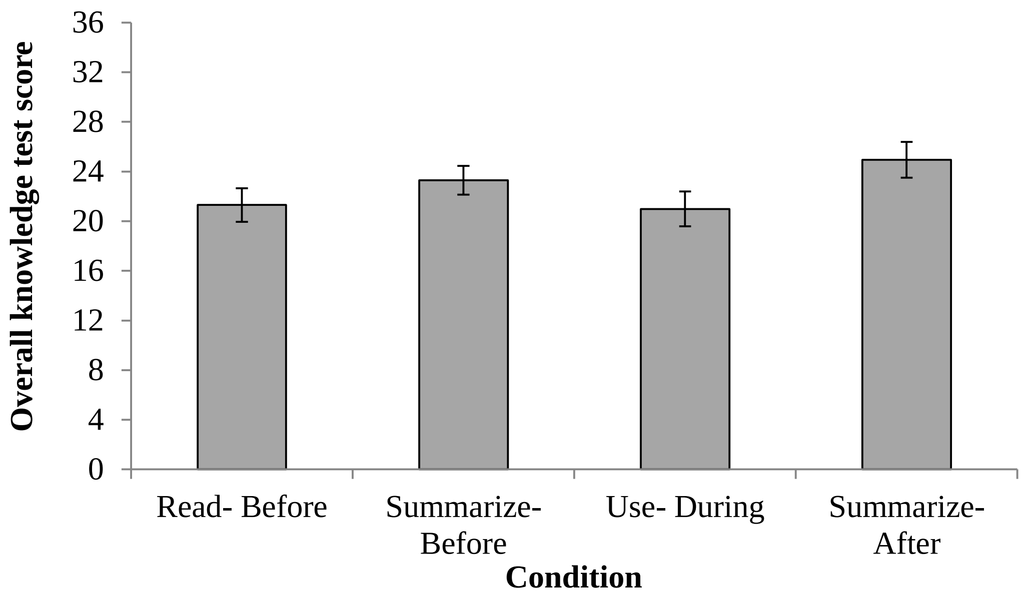


Figure 5. Average knowledge test scores (on the scale from 0 to 36) for each of the four conditions. The error bars show twice the standard error, approximating the 95% CI (Cumming & Finch, 2005).

The hypothesis was partly supported: An overall ANOVA showed a significant effect of condition on the knowledge test score ($F(2, 128) = 8.82$, $p < .001$, $\eta_p^2 = .12$), and a post-hoc analysis using Tukey's HSD showed that the participants in the Summarize-Before and Summarize-After groups scored significantly higher on the knowledge tests compared to the participants in the Use-During group ($p < .05$ in both cases). However, there was no difference found between the scores of the Summarize-Before and Summarize-After groups ($p > .05$), and the second part of the hypothesis was not supported.

Method of Studying the Principles (Hypothesis 2)

The second hypothesis predicted that the participants who read the principles before training (Read-Before) would have worse knowledge of the system compared to participants who summarized the principles before training (Summarize-Before). This hypothesis was supported: The participants in the Read-Before condition had significantly lower score on the knowledge tests compared to participants in the Summarize-Before condition, $t(89) = -2.22$, $p < .05$, $d = 0.57$, 95% CI [-3.77,-.211]. The average difference was only two questions on average (see Figure 5).

Does the Principle Study Time Affect Knowledge Test Outcome?

Because participants spent different amount of time studying the principles I wanted to investigate whether more time used to study the principles would deliver proportionally better learning, especially given the small difference in scores among the groups. To do this, I divided the total time taken to study the principles with the number of correct questions for each participant. Thereby calculating the average study time needed for every correct answer.

There was an overall difference among the groups, $F(3, 171) = 53.52$, , $p < .001$, $\eta_p^2 = .48$, and a post-hoc comparison with Tukey's HSD showed that there was a significant difference between all groups ($p < .05$) except Summarize-Before and Summarize-After ($p > .05$; see Figure 6). The participants in the Use-During group spent significantly less time studying the principles for every correct answer on the knowledge test compared to all the other groups, and the participants in the Read-Before group spent significantly less time studying the principles for every correct answer on the knowledge test compared to both summarizing groups. Therefore, for every correct answer on the knowledge test the participants summarizing the principles needed considerable more time (almost twice as long) to study the principles. Even if the Summarize-Before group scored higher on the tests, this came at the cost of considerably longer study time for each correct item.

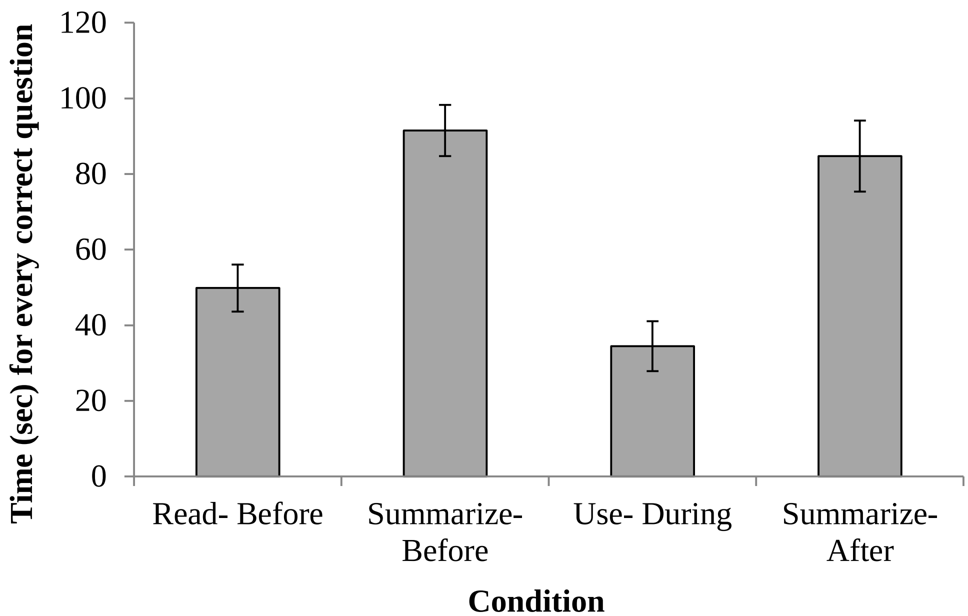


Figure 6. Time spent studying the principles (in seconds) for every correct question on the knowledge tests. The error bars show twice the standard error.

A hierarchical regression was used to investigate whether condition influenced knowledge test outcome over and above the time spent studying the principles. Study time was entered in the first step and condition (dummy coded) was entered in the second step. The time studying principles was a significant predictor of knowledge test score and accounted for 13% of the variance (see Table 6). When the time spent studying the principles had been accounted for, condition did not predict knowledge test outcome.

This indicates that the difference among the conditions is mostly due to the difference in the time spent studying the principles.

Table 6
Summary of the hierarchical multiple regression analysis predicting knowledge test outcome based on the time studying the principles and condition (dummy coded where the Read-Before group served as comparison).

Predictor	R^2	Standardized Coefficient (β)
Step 1	$.13^{**}$	
Time spent studying the principles		0.36^{**}
Step 2	.17	
Summarize-Before		-0.12
Use-During		0.03
Summarize-After		0.12

$^{*}p < .05$. $^{**}p < .001$.

Does Prior Knowledge Predict Knowledge Test Outcome?

What participants already knew about electrical circuits could be expected to influence the outcome on the knowledge tests. Prior knowledge was represented by three variables: prior experience, self-reported domain knowledge, and a test of prior knowledge. The prior knowledge variables were entered as predictors in the first step of a hierarchical multiple regression (the categorical variables were dummy coded) and condition was entered in the second step. The regression showed that prior knowledge significantly predicted knowledge test outcome ($R^2 = .19$, $F(4,172) = 10.30$, $p < .001$), but condition also significantly predicted knowledge test outcome when prior knowledge had been statistically accounted for ($R^2 = .31$, $\Delta R^2 = .12$, $F(7,169) = 10.59$, $p < .001$).

Conclusions on Knowledge of the System (Hypotheses 1 & 2)

The results indicate that knowledge of the system is influenced by how the principles are studied: The participants who summarized the principles scored higher on the tests of knowledge of the system than the participants who used the instructions during training (hypothesis 1), and those who read the principles beforehand (hypothesis 2), respectively.

The benefit in terms of number of correct answers was not large however, and when time needed to study the principles for every correct answer was calculated, the participants in the Use-During and Read-Before conditions needed significantly less study time per correct answer compared to the participants who summarized the principles. These results demonstrate a trade-off between score and study time; summarizing yields a higher absolute score on a test but at the cost of substantially longer study time per correct item, and lower instructional efficiency. Therefore, the choice of requiring learners to employ a certain study method must be made with the fact in mind that this can come at a cost of instructional efficiency.

When time taken to study the principles had been statistically accounted for condition was not a predictor of knowledge test outcome, indicating that the crucial factor here is the study time. Taken together with the results of lower instructional efficiency for the summarizing groups, it seems the time studying the principles is important, but requiring learners to summarize all the information might not be the most efficient use of their time.

The first hypothesis also predicted that the participants in the Summarize-Before condition would score higher on the knowledge tests than the participants in the

Summarize-After condition. This part of the hypothesis was not supported and the results did not support the idea that the principles would act as advance organizer and enhance knowledge of the system for the participants in the Summarize-Before condition.

Prior domain knowledge was associated with better knowledge test outcome, but when prior knowledge had been statistically controlled, condition uniquely accounted for variance of the knowledge test scores. Therefore, differences in prior knowledge do not fully explain the results.

Procedural Learning Measured with Troubleshooting Performance (Hypotheses 3 & 4)

Procedural learning was measured by having participants complete troubleshooting tasks without the aid of instructions and a number of different dependent measures were used: The time needed to complete each testing task, the number of safety errors made (requiring a restart of the task), number of unnecessary components replaced, and number of meter readings (ohmmeter and voltmeter). Each of the dependent variables was standardized; the group mean was subtracted from an individual score and the result divided by the standard deviation, creating a score in terms of standard deviations from the mean, which is represented by zero. The standardized scores of the four measures were then averaged to create a single task performance score.

Levene's test of homogeneity of variance was calculated to guarantee ANOVAs would be the appropriate statistics to use with the independent variable having unequal sample sizes. The test was not significant for any of the measures ($p > .05$ in all cases) and the assumption of equal variance holds.

Timing of Principle Use (Hypothesis 3)

The third hypothesis stated that participants in the Use-During group were going to show better procedural learning (as measured on the troubleshooting tasks) than participants in the Summarize-Before and Summarize-After groups. In addition, I expected that the participants in the Summarize-Before group would have better procedural learning than the participants in the Summarize-After group.

The average standardized performance measure for all four conditions is shown in Figure 7 (all four groups are shown here for completeness, even if the current hypothesis only addresses the Summarize-Before, Use-During and Summarize-After conditions). A positive standardized measure indicates performance that is worse than the average, as participants would have taken longer to complete the testing tasks, making more safety errors, unnecessarily replaced more components, and used more meter readings in the process. Conversely, a negative standardized measure indicates performance that is better than the average.

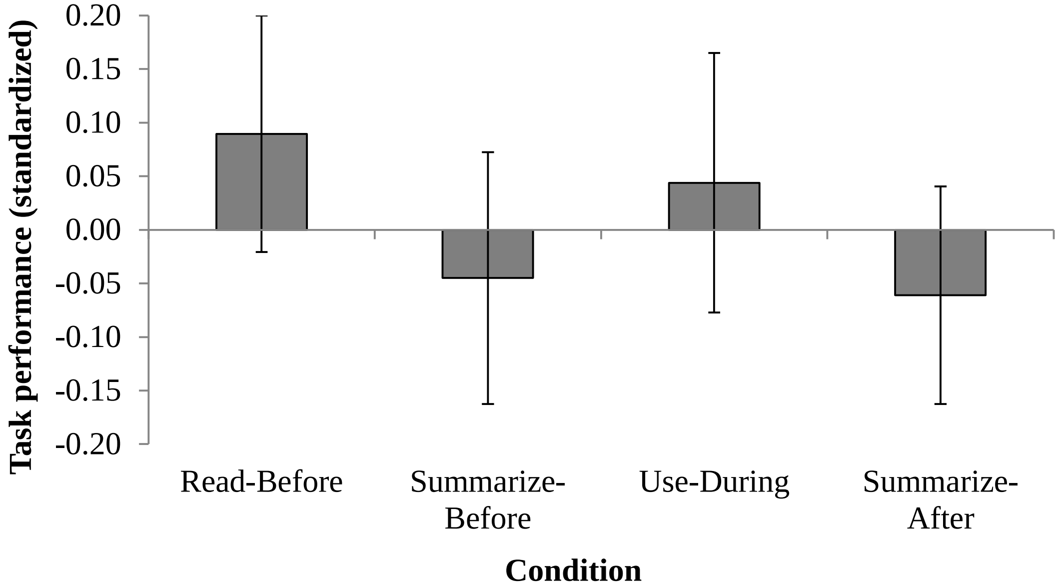


Figure 7. The average standardized performance measure of the testing tasks for each condition. A positive score indicates performance worse than the average, whereas a negative score indicates performance better than the average. The error bars show twice the standard error.

There was no significant difference among the groups (p > .05). The hypothesis was therefore not supported, and in fact, the trend of the averages was in an opposite direction to what had been expected as the summarizing conditions had better outcome on the testing tasks than the Use-During condition, and the Summarize-After condition had a better outcome than the Summarize-Before condition.

<u>Method of Studying the Principles (Hypothesis 4)</u>

The fourth hypothesis was aimed at examining how the method of studying affects procedural learning. I hypothesized that the participants in the Summarize-Before group would have better knowledge of the system and procedural learning than the participants in the Read-Before group.

The participants in the Summarize-Before condition generally performed better on the testing tasks compared to the participants in the Read-Before condition, but this difference was not significant ($p > .05$).

Type of Testing Task

Generally the participants performed worse when completing the transfer tasks ($M = 0.04$, $SD = 0.65$) than the retention tasks ($M = -0.13$, $SD = 0.69$), and this difference was significant ($F(1, 630) = 8.76$, $p < .05$, $\eta_p^2 = .01$). There was no interaction between condition and test type (retention vs. transfer; $p > .05$).

Hints To Complete Testing Tasks

When working on the testing tasks, the participants had access to hints to help them complete the tasks. Once a certain amount of time had elapsed (15-20 minutes depending on the task), the first hint appeared and then the other two three and six minutes later respectively. The first hint revealed what type of fault was involved (open or short) and whether the voltmeter or the ohmmeter should be used. The second revealed the area of the circuit where the fault was located, and the third hint explained exactly what the fault was and what had to be done to fix it. Therefore, if participants used the third hint they were considered unsuccessful at completing the task on their own.

Figure 8 shows the percentage of testing tasks (each participant completed seven testing tasks) where hints were used. Generally, the participants in the Use-During and Read-Before groups used hints more than the participants in the summarizing groups but there was no difference among the groups for any of the three hints ($p > .05$ in all cases).

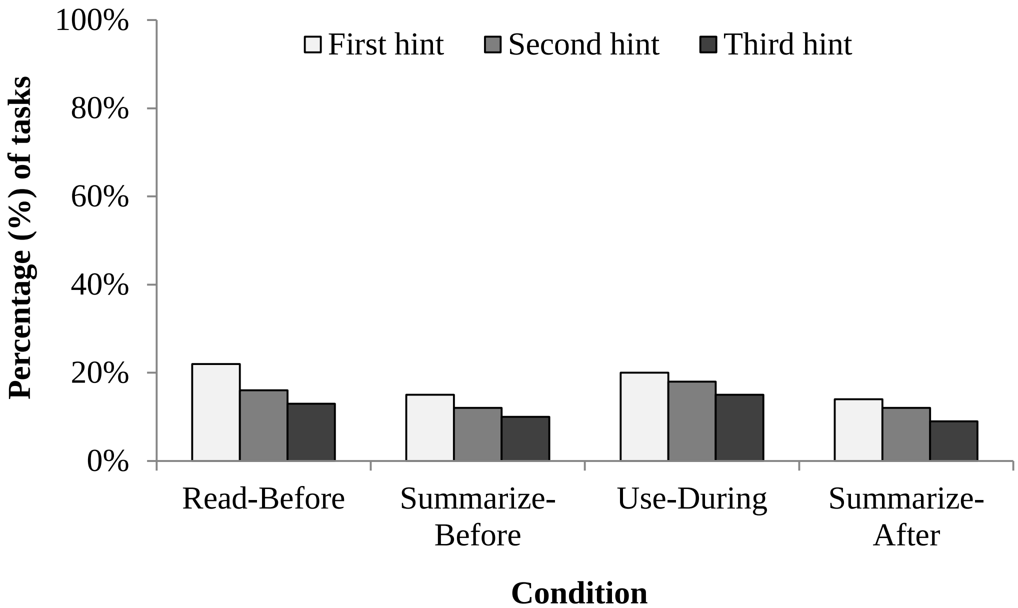


Figure 8. Percentage of participants in each condition who used the hints to complete the testing tasks.

Does Knowledge of the System Predict Procedural Learning?

The results on the comparison of how the principles are studied showed no difference between the Read-Before and Summarize-Before groups on any measure of procedural learning. This was unexpected given that the participants in the Summarize-Before condition had shown an advantage when measured on the knowledge of the system. This raises the question to what degree the knowledge of the system translates into better procedural learning. To get at this, I ran a multiple regression for the measure of procedural learning using score on the knowledge tests as a predictor.

The average score on the knowledge tests significantly predicted the procedural learning outcome (R^2 = .06, $F(1,624) = 42.76$, $p < .001$, β = -0.25). The higher the score on the knowledge tests, the better the procedural learning outcome.

Conclusions on Procedural Learning (Hypotheses 3 & 4)

In the third hypothesis I predicted that the timing of principle use (before, during, or after training) would influence procedural learning. That the participants using the principles while completing the training tasks would show better procedural learning, as measured with troubleshooting performance, than the participants who either summarized the principle before or after training. In addition, I had expected the participants who summarized the principles before training would show better procedural learning than the participants who summarized the principles after training. In the fourth hypothesis I predicted that the participants who studied the principles by summarizing them would show better procedural learning than participants who read the principles.

The evidence did not support any of my predictions: The measures of procedural learning did not show any differences among the groups and there was no difference among them in terms of whether participants needed hints to help them complete the testing tasks. However, participants performed worse when completing transfer tasks than retention tasks and this was true for all the conditions. This indicates that the transfer tasks were generally considered more difficult than the retention tasks.

Overall, there was little effect of timing or study method used to process the principles on procedural learning, indicating that these variations in providing principles are not important for determining procedural learning. Knowledge of the system was however a significant predictor of procedural learning (albeit explaining only a small proportion of the variability), indicating that it does play a role; it just might be a minor one.

One could hypothesize that the lack of effect might be due to the measures of procedural learning not being sensitive enough to measure the difference, but I would argue against this interpretation. First, I did find difference between retention and transfer tasks, which is to be expected when comparing these two different types of tasks. Therefore, there is some evidence that the measure is sensitive enough. Second, a number of different variables were used to create the measure of procedural learning (time-on-task, number of safety errors, number of unnecessarily replaced components, and number of meter readings) and each can be considered a theoretically valid measure of troubleshooting performance in this task domain. It is difficult to imagine what operationalization of procedural learning would serve better in this context, especially as the simulation is created by subject matter experts and these are the measures they use to evaluate troubleshooting training.

Training Performance (Hypothesis 5)

Training performance was measured in the same way as performance on the testing tasks, with time-on-task, number of safety errors, unnecessarily replaced components, and meter readings. These measures were all standardized and then averaged for a composite standardized measure of training task performance.

In the fifth hypothesis I proposed that the order of training performance from best to worst would be: Use-During, Summarize-Before, Read-Before, and Summarize-After. I expected differences in performance between the lasts two groups because I expected some participants in the Read-Before group to study the principles with some effort and therefore increase the average performance for that group.

The average standardized performance for the conditions on the three training tasks is shown in Figure 9. Generally, the Summarize-Before condition had the best performance, but there was no significant difference among the groups ($p < .05$). The hypothesis was therefore not supported as there was no difference in training performance among the four groups. The evidence therefore does not indicate that the participants with access to principles during problem solving would benefit from problem solving with principles to guide them.

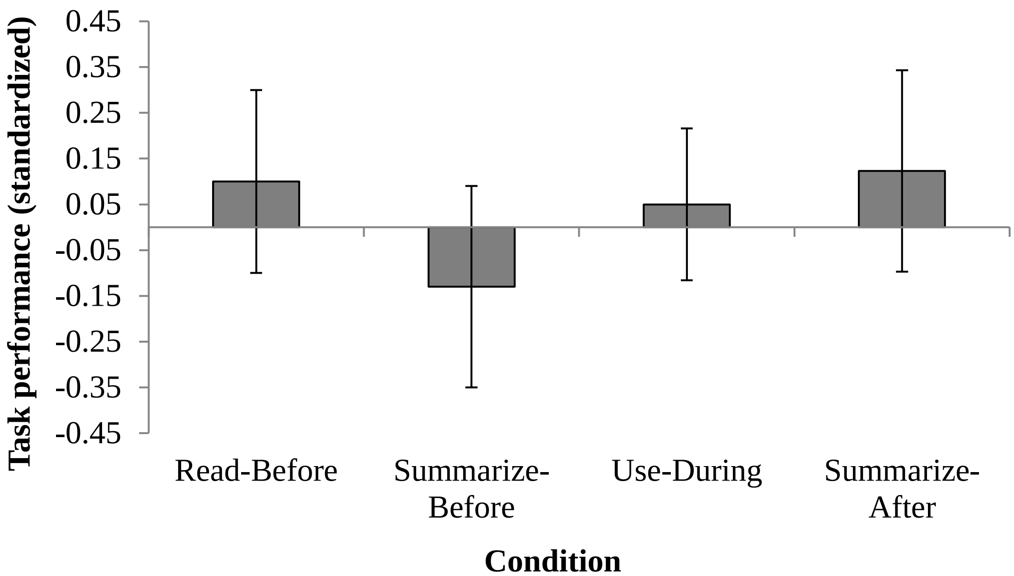


Figure 9. Averaged standardized performance on the three training tasks for each condition. The error bars show twice the standard error.

Subjective Workload (Hypothesis 6)

Subjective workload was measured with an abbreviated NASA-TLX after each task (both training and testing). The NASA-TLX measures workload on six different dimensions of which five were used (mental demand, temporal demand, success at task, difficulty of obtaining that level of success, and degree of frustration). The sixth dimension represents physical demand, but was excluded as all the tasks were carried out on a computer and did not involve any differential physical demand. Each dimension was rated on the scale from 0 to100. The score for each of the five scales were averaged to create a single score for general subjective workload (after the dimension for success was reversed), where a lower number represented less subjective workload.

At the beginning of the experiment, before the participants started the training, they were given a simple example task to familiarize them with the simulation. After completing the example task (using detailed procedural instructions telling them exactly what to do), the participants rated the subjective workload of the task. They were explicitly informed that the example task would be the easiest task in the experiment and to bear this in mind when rating the task. This was done to provide a baseline against which to compare all other ratings and to provide a way to account for individual differences in the use of the rating scales. Therefore subjective workload ratings will not only be considered in absolute terms, but also in relation to the baseline rating of the example task.

In the sixth hypothesis I proposed that the reported subjective workload would be in the following order during training (from lowest to highest): Use-During, Summarize-Before, Read-Before, and Summarize-After. But during testing, the Summarize-After group would report lower subjective workload than the Read-Before group because in the meantime they had summarized the principles and could use this information to help solving the tasks.

Overall the participant's average reported workload was between 30 and 50 (out of 100). The participants in the Read-Before condition reported the highest workload and the participants in the Summarize-After reported the lowest, both during training and testing (see Figure 10).

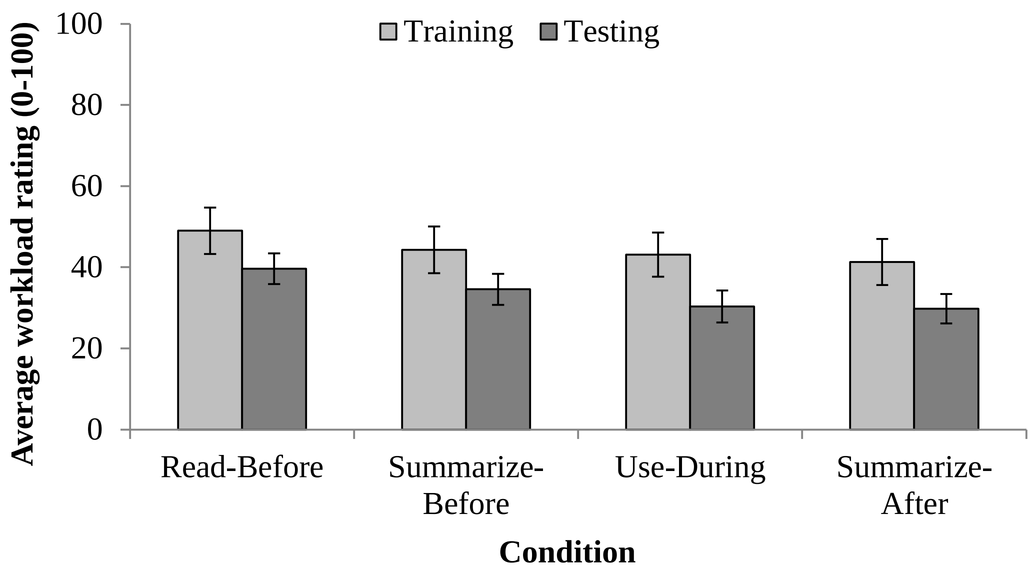


Figure 10. Average subjective workload rating for training and testing and each condition. The error bars show twice the standard error.

The difference among the conditions was not significant for the training task ratings ($p > .05$), but it was for the testing task ratings ($F(3,627) = 5.82$, $p < .001$, $\eta_p^2 = .03$). A post-hoc comparison with Tukey's HSD showed that during testing the Read-Before group reported a significantly higher workload than the Use-During and Summarize-After groups respectively ($p < .05$ in both cases). This is to a degree similar to what was predicted, as the participants in the Read-Before condition reported the highest workload in testing, but contrary to what was predicted, there was no difference in reported workload between the Summarize-After and Use-During conditions.

The participants reported a higher average workload for the training tasks than they did for the testing tasks ($t(906) = 6.25$, $p < .001$, $d = 0.42$, 95% CI [7.40, 14.17]). In addition, the participants reported a higher average workload for the transfer tasks ($M = 37.46$, $SD = 24.16$) than the retention tasks ($M = 28.69$, $SD = 23.18$) during testing ($t(629) = -4.59$, $p < .001$, $d = -0.37$, 95% CI [-12.51, -5.02]).

Subjective Workload with Baseline

These kinds of subjective measures of workload can be criticized on the grounds that participants differ in how liberal or conservative they are when using the rating scales. For example, some participants might never use the highest possible rating whereas others consistently do so. It is, in other words, difficult to determine whether the variance seen is due to difference in task workload (as a function of condition) or individual differences in how the rating scales are used.

One way to counteract this is to have all participants complete a single task in the same way and ask them to rate it. The variance in the rating should reveal individual differences as the task difficulty is kept constant. This was the rationale for asking all participants to rate the example task at the beginning of the experiment.

To take the baseline ratings into account the workload ratings for the example tasks were first averaged into a single workload score (average baseline rating) for each participant and then the average baseline rating was subtracted from the average workload score for each task and participant (i.e., participant SB-10 had average baseline rating of 4.4 and so this number was subtracted from the rating of each task this participant completed). The resulting variable represents the workload rating over and above the baseline rating for each participant, therefore taking into account individual differences in how liberal or conservative participants might be in rating the workload.

As was seen with the absolute workload ratings, there was no difference among the groups in training ($p > .05$). But there was a significant difference among the groups for the testing tasks ($F(3,627) = 3.42$, $p < .05$, $\eta_p^2 = .02$). The Read-Before group reported a significantly higher subjective workload when completing the testing tasks compared to the Summarize-After group ($p < .05$; see Figure 11).

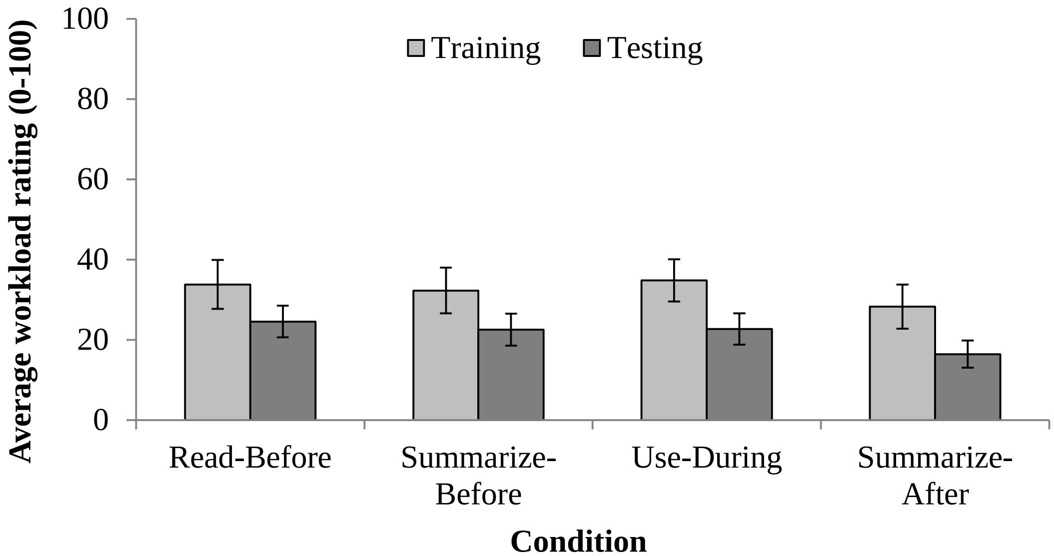


Figure 11. Average workload rating when baseline has been taken into account for the training and testing tasks and each condition. The error bars show twice the standard error.

When the baseline ratings of subjective workload had been taken into account the participants generally reported more workload during training ($M = 32.27$, $SD = 23.64$) than testing ($M = 21.59$, $SD = 24.20$; $t(906) = 6.17$, $p < .001$, $d = 0.41$, 95% CI [7.28, 14.08]) and higher workload for transfer tasks ($M = 25.30$, $SD = 24.26$) as compared to retention tasks (*M* = 16.66, *SD* = 23.26; $t(629) = -4.51$, $p < .001$, $d = -0.36$, 95% CI [-12.41, -4.88]).

Conclusions on Subjective Workload (Hypothesis 6)

In the sixth hypothesis I predicted that ratings of workload would be lowest for the Use-During group, and then the Summarize-Before group, both during training and testing. In addition, I had predicted that the Summarize-After group would rate workload and task difficulty higher than the other groups during training (because they only had access to general procedural instructions to help them), and the Read-Before group would do so during testing (because by then the participants in the Summarize-After group would have studied the principles and this would aid them in the testing tasks, whereas the Read-Before participants would be less able to rely on the principles).

There was no difference in the subjective workload during training, but the participants in the Read-Before condition reported higher workload than the participants in the Use-During and Summarize-After conditions during testing. This pattern is in line with what had been expected for the testing tasks, in terms of the Read-Before condition reporting the highest subjective workload, but the hypothesis was not supported in any other way.

After baseline workload ratings had been taken into account, there still was no difference among the conditions in training, and only the difference between the Summarize-After and Read-Before groups remained for the testing tasks. Taking the baseline rating of subjective workload into account provides a more conservative way of considering workload than just looking at absolute scores, as individual variances in the use of the rating scales have been minimized. Therefore, the difference in reported workload between the Summarize-After and Read-Before groups in testing is most likely due to the manipulation and not any other factors. An explanation for this finding is that in testing the participants in the Summarize-After group had recently studied the principles and could use them to understand what they had been doing right or wrong while completing the training tasks. That is, it gave them an opportunity to reflect on their performance and try to learn from their mistakes. This might have given them some advantage when doing the testing tasks, at least when compared to participants who read the principles before doing the training tasks.

The lack of difference in subjective workload among the groups during training is surprising, especially because the participants in the Summarize-After group had at the point not had any exposure to the principles. It is impossible to say anything definite about non-significant results, but a tentative suggestion could be that the participants in the Summarize-After group reported less workload than expected because they did not have to integrate multiple sources of information while completing the training tasks, whereas having to do so might have added to the experienced workload of the other three groups. This might have counteracted the proposed benefit from having more information available to assist in problem solving. However, the observed power of the analysis was not high enough (only 0.08 when 0.8 is considered adequate power) to state that the null results were due to an absence of effect and not lack of power.

There was a significant difference in subjective workload between training tasks and testing tasks, both overall and when baseline ratings had been taken into account, with more workload being reported during training as compared to testing. In addition, participants generally reported more workload for transfer tasks than retention tasks, and this was true both overall and when baseline ratings had been taken into account.

Instructional Use

During training the time spent viewing each piece of instruction (principles or procedural information) was measured in seconds. The access to principles varied according to condition but all participants had access to procedural instructions during training task completion.

I expected that the participants in the Summarize-Before and Summarize-After groups would spend more time viewing the principles than the Read-Before and Use-During groups because of the nature of the manipulation (having to summarize each principle). The duration of viewing each piece of procedural instructions will be compared in the same manner as the principles, but I did not predict any specific pattern of results in this case. Figure 12 shows the average time used to study principles and procedural instructions for each group.

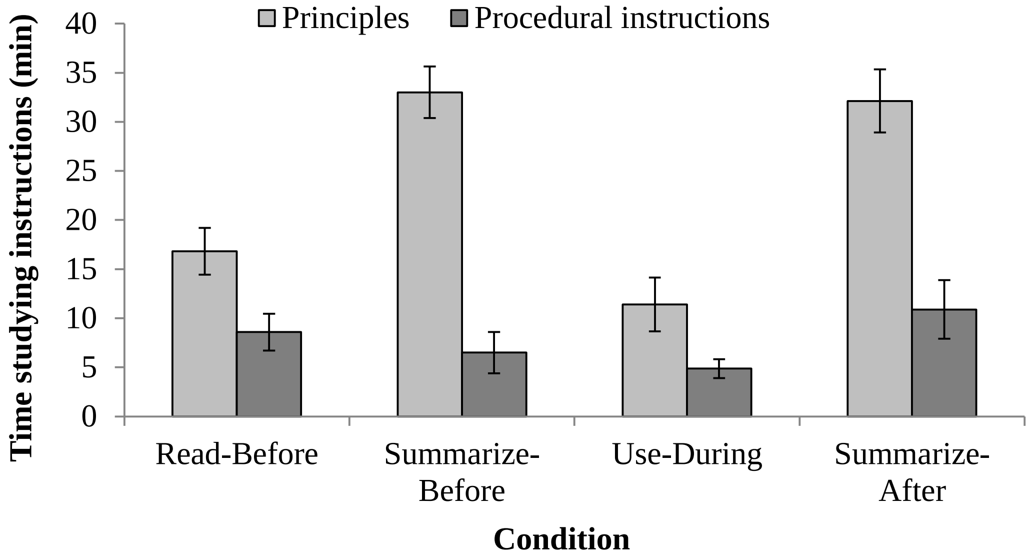


Figure 12. Time spent studying principles and procedural instructions in minutes for each condition. The error bars show twice the standard error.

Principle Use

It is clear that the participants in the two summarizing condition spent considerably more time studying the principles as compared to the other two groups, and the overall difference among the groups was significant ($F(3,86) = 61.44$, $p < .001$, $\eta_p^2 = .68$) as well as all pairwise comparisons (with Tukey's HSD, $p < .05$ in all cases), except the comparison between Summarize-Before and Summarize-After ($p > .05$). The hypothesis was therefore supported as the participants in the summarizing groups spent more time studying the principles compared to the other two groups. The difference was quite large, with 15-20 minutes difference between the summarizing conditions and the other two.

The Read-Before participants spent significantly more time (on average 5 minutes) studying the principles than the Use-During participants, indicating that however the participants in the Use-During group were studying the principles, they spent less time doing so than the participants who read the principles.

Because of the difference among the conditions found for the time spent studying the principles, I wanted to investigate whether the time spent studying the principles would predict procedural learning. A multiple regression showed that the time spent studying the principles did not predict procedural learning (standardized performance on the testing tasks), $R^2 = .01$, $p > .05$. Also, including condition did not change the variance explained by the regression equation.

<u>Procedural Instructions Use</u>

There was a significant overall difference among the conditions for the time spent studying the procedural instructions, $F(3,89) = 6.11$, $p < .001$, $\eta_p^2 = .17$. A follow-up with Tukey's HSD showed that the participants in the Summarize-After spent significantly more time studying the procedural instructions compared to the Summarize-Before and Use-During groups respectively ($p < .05$ in both cases).

Because of the difference among the conditions, I used a multiple regression to investigate whether the time spent studying the procedural instructions predicted procedural learning. The time spent studying the procedural instructions was a significant predictor of procedural learning (averaged standardized measure of performance on the testing tasks), $R^2 = .03$, $F(1,630) = 17.15$, $p < .001$, $\beta = 0.16$. The longer participants spent studying the procedural instructions, the poorer their performance on the testing tasks. These results indicate that participants who studied the procedural instructions did not learn more about how to troubleshoot the circuit. Instead, participants who found it difficult to complete the training tasks and understand what they needed to do probably studied the procedural instructions for a longer time.

Conclusions on Instructional Use

Because of the experimental manipulation, I had predicted that the participants in the summarizing conditions would spend more time studying the principles than the participants in the other two conditions, and this hypothesis was supported. The difference was quite large when comparing the participants summarizing the principles and using them during task completion (15-20 minutes), but those who read the principles also spent more time studying the principles (about 5 minutes longer) than participants who only used the principles during training. Condition explained a large proportion of the variance (68%). Providing participants with a particular study method, especially asking them to summarize the principles, led them to devote more time to study the principles than if they were not asked to study them in a particular manner. However, the time spent studying the principles did not predict testing task performance.

The participants in the Summarize-After condition spent significantly more time studying the procedural instructions as compared to the Summarize-Before and Use-During groups. This was most likely because this was their only source of information when completing the training tasks, and had no other information to rely on when figuring out what they needed to do. The time spent studying the procedural instructions did predict procedural learning but in an unexpected way: Longer study time was associated with poorer performance. I interpret this to suggest that as participants had more trouble with completing the training tasks they studied the procedural instructions more.

Changes over Time – Difference between Testing Sessions

Participants were tested on two different occasions: immediately after training (immediate testing) and a week later (delayed testing). I predicted similar pattern of results for both testing occasions in terms of the effects of condition (i.e., no interaction between condition and testing occasion), but I expected generally better learning outcomes on the immediate compared to the delayed testing session. This would be due to the differences between long-term and transient effects of training, as the immediate testing session is closer in time to the learning episode.

On the tests of knowledge of the system there was not a significant difference between the immediate (M = 22.63, SD = 4.65) and the delayed (M = 22.57, SD = 4.78) testing occasion (p > .05). On procedural learning there was also not a significant difference between the immediate (M = -0.05, SD = 0.67) and the delayed (M = 0.06, SD = 0.75) testing sessions (p > .05). The hypothesis was therefore not supported.

I had expected that participants would generally do worse on the delayed than the immediate session, because on the immediate session transient effects of the training would help performance on the testing tasks, but after a week these transient effects would have disappeared. The results did not support this hypothesis.

Summaries Created by the Participants

The participants in the Summarize-Before and Summarize-After groups were required to study the principles by writing a summary from each one. These summaries were coded for completeness (content points) and to what degree they were rephrased by the participants (summary points). The assumption was that as the summaries are more complete (covering more of the content in the principles) and more fully rephrased by the participant (less verbatim) the better the participants would learn the information, regardless of whether they studied the principles before or after the training tasks. In both cases a higher score represented a better outcome: More content points meant the participants included more content in the summaries (the maximum score was 43 points), and more summary points meant the participants were more likely to use their own words in summarizing the content (the maximum score was 22 points).

Multiple regressions with average knowledge test score and the procedural learning measure as the dependent variables and the content and summary points as predictors, showed that summary points predicted outcome on knowledge tests and procedural learning outcome (see Table 7). More summary points (meaning less verbatim the summaries) predicted higher score on the knowledge tests and better performance. Content points did not predict outcome on knowledge tests or procedural learning measures.

Table 7
Summary of the multiple regression analysis predicting the knowledge test and procedural learning outcomes from the content and quality of the summaries.

Predictor	Knowledge test score (β)	Procedural learning (β)
Content points	0.21	0.04
Summary points	0.30^{*}	-0.20^{**}
R^2	$.11^{*}$	$.04^{**}$
F	5.03^{*}	7.06^{**}

$^{*}p < .05$. $^{**}p < .001$.

The degree to which participants in the summarizing conditions rephrased the principles using their own words while summarizing the content predicted both outcome on the knowledge tests and performance on the testing tasks. The less verbatim the summaries the higher the participants scored on the knowledge test and the better their performance on the testing tasks. The completeness of the summaries, that is, how much of the content of the principles they covered, was not found to predict either learning outcome. Taken together this provides an interesting insight into how summarizing might work as a study method. It is more important for learning to have learners make an effort to put the material to-be-learned in their own words, than to have them cover all the content in the material.

When the two conditions, Summarize-Before and Summarize-After, were compared there was no difference between the groups in terms of summary points (Summarize-Before: M = 19.48, SD = 3.35; Summarize-After: M = 19.35, SD = 3.45), but there was a significant difference in the number of content points, $t(90) = 2.97$, $p < .05$, $d = 0.63$, 95% CI [1.23,6.17]. The Summarize-Before group had fewer content points (M = 27.09, SD = 6.12) than the Summarize-After group (M = 30.78, SD = 5.80). This means that the participants in the Summarize-After group generally summarized more of the content of the principles than the participants in the Summarize-Before group (but note that there was no difference between the groups in terms of time needed to summarize the principles). One might wonder whether this is because the participants in the Summarize-After condition realized how this information might be applied to the tasks, whereas the participants in the Summarize-Before condition did not.

Individual Differences

A multiple regression was used to investigate the effects of individual difference variables on learning outcomes (knowledge of the system and procedural learning). Individual difference variables were demographic information (gender and age), GPA, SAT scores (verbal and math), and preferred learning modality as measured on the VARK learning styles questionnaire (resulting in four scores representing preferences for visual, aural, read/write, and kinesthetic modality). These variables were entered as predictors into two separate analyses of knowledge of the system and procedural learning.

Table 8 summarizes the regression results for the two learning outcome variables. The significant predictors for knowledge test score were age, and the visual and kinesthetic scores on the VARK questionnaire. Older participants scored higher on the knowledge tests, and so did those showing preference for kinesthetic learning. The direction of the association was reversed for the participants preferring visual modality; they tended to score lower on the knowledge tests. None of the variables entered significantly predicted procedural learning.

Table 8
Summary of the multiple regression analysis with individual difference variables predicting knowledge of the system and procedural learning.

Predictor	Knowledge test score (β)	Procedural learning (β)
Gender	0.16	-0.11
Age	0.29**	-0.03
GPA	0.15	-0.05
SAT-Verbal	0.11	0.05
SAT-Math	0.15	-0.07
Visual score	-0.24*	0.001
Aural score	-0.05	0.07
Read/write score	-0.01	-0.03
Kinesthetic score	0.20*	0.001
R^2	.23**	.02
F	4.06**	1.23

*$p < .05$. **$p < .001$.

It is difficult to say why older participants tended to have better outcome on the knowledge tests; some reasons could for example be greater exposure to circuit information or more conscientiousness.

Kinesthetic modality refers to "perceptual preference related to the use of experience and practice (simulated or real)" (Fleming & Mills, 1992, p. 140-141). Therefore, participants who showed preference for learning through experience generally scored higher on the knowledge test. It can be hypothesized that this was because these participants benefited more from doing the training tasks (which provides a hands-on experience), in terms of gaining understanding of the system, compared to participants who do not prefer learning kinesthetically.

A preference of a visual modality involves preferring to learn from depicted information, such as graphs, diagrams, or symbols. That is, visual but non-verbal information (Fleming & Mills, 1992). A higher preference for visual modality was associated with lower scores on the knowledge tests, and it can be hypothesized that as the instructional materials and tests of knowledge of the system were primarily verbal and those preferring to learn using visual but non-verbal material suffered as a consequence.

Drawings and Explanations of the Circuit

The drawings and explanation of the circuit that the participants were asked to provide at the end of the delayed testing session were included to attempt to elicit the participant's mental model of the system.

Drawings of the Circuit

The drawings were coded according to three criteria: whether they depicted the structure of the circuit, whether they provided details explaining the function of the circuit, and whether they generally showed details. The scores were generally high for all four conditions (see Table 9), indicating that the participants had a good memory for the structure of the circuit and details important for the functioning of the circuit. The participants did not score as high on the drawing details, suggesting they were less likely to take the time to draw functionally irrelevant details. There was no difference among the conditions found for the drawing scores ($p > .05$ in all three cases).

Table 9
Means and standard deviations (in parentheses) on each component of the drawings (structure, functional details, and drawing details) for every condition. The possible scores for each component are provided in parentheses in the column headings.

Condition	Structure points (0-26)	Functional detail points (0-7)	Drawing detail points (0-6)
Read-Before	24.83 (0.92)	5.06 (1.47)	2.73 (0.97)
Summarize-Before	24.68 (1.17)	4.18 (1.96)	2.70 (0.97)
Use-During	25.00 (0.97)	4.40 (1.56)	2.78 (1.03)
Summarize-After	25.33 (0.73)	5.14 (1.64)	2.98 (1.21)

Explaining the Circuit

The participants answered three explanation questions. First, they were asked to provide a short explanation of all the components of the circuit, using a picture of the circuit for reference. Second, they were asked to describe the workings of the relay in detail, again using a picture of the relay for reference. Third, they were asked to mark on a picture of the circuit where it would be energized when the lights are off.

On the first question, there was no difference among the groups in how well they explained the functioning of the components in the circuit ($p > .05$; see Table 10). On the second question (explaining the relay in detail), it is notable that the scores were overall very low, and most participants had a difficult time explaining the function of the relay in detail. There was a significant difference among the groups ($F(3,83) = 2.76$, $p < .05$, $\eta_p^2 = .09$), and Tukey's HSD post-hoc tests showed that the Summarize-After group had a higher score on average than the Summarize-Before group ($p < .05$). For the third question, participants were asked to indicate where the circuit would be energized, given that everything functioned normally and the lights were off. Again, the Summarize-After group had the highest scores, and the participants in the Read-Before group had the lowest, and there was a significant difference only between these two groups ($F(3,83) = 2.93$, $p < .05$, $\eta_p^2 = .10$, and $p < .05$ with Tukey's HSD post-hoc comparison).

Table 10

Means and standard deviations (in parentheses) for each explaining question and every condition. The possible scores for each question are provided in parentheses in the column headings.

Condition	Question1: Explain function of components (0-27)	Question 2: Explain how the relay works in detail (0-22)	Question 3: Depict where circuit is energized (0-7)
Read-Before	7.30 (4.18)	1.30 (1.26)	1.37 (1.72)
Summarize-Before	6.87 (3.49)	0.83 (0.72)	2.04 (2.31)
Use-During	5.55 (3.36)	1.63 (1.44)	1.93 (2.13)
Summarize-After	7.29 (3.00)	1.86 (1.53)	3.26 (2.45)

Conclusions on the Drawings and Explanations of the Circuit

There was no difference among the conditions in terms of their drawings of the circuit, regardless of whether the drawings were scored based on structural content, functional details depicted or general details provided in the drawings.

There was no difference among the groups in terms of how well they were able to explain the functioning of the components in the circuit, but the Summarize-After group explained the functioning of the relay better than the Summarize-Before group and was more accurate when indicating where the circuit would be energized than the Read-Before group. It is interesting that the Summarize-After group is showing better outcome in these cases, especially in comparison to the Summarize-Before group which used the same study method for the principles. One can only assume that having completed the training tasks when they studied the principles helped these participants put the information in the principles in context and create a better mental model of the circuit.

Discussion

The primary purpose of the first experiment was to understand how the timing of principles use would influence procedural learning and knowledge of the system. I had predicted that different factors would influence these two categories of learning outcomes differently: That knowledge of the system would increase when participants summarized the principles either before or after completing training tasks, but the knowledge of the system would suffer when the principles were used in the context of task completion. I expected the opposite for procedural learning: Using the principles during task completion would lead to better procedural learning compared to studying the principles outside of doing the training tasks.

These expectations were based on predictions about how the principles would be processed. Studying them explicitly by summarizing them was expected to be beneficial for knowledge of the system, but using them in the context of doing the procedural tasks would be beneficial for procedural learning. All the participants received general procedural instructions in training (which did not provide detailed information on each step of the procedure) and therefore required the participants to engage in some problem solving. I believed that in the context of problem solving the principles would become important.

The first part of the prediction was supported (Hypothesis 1); participants who summarized the principles scored higher on tests of system knowledge than participants who used the principles during training task completion. The second part of the prediction was not supported (Hypothesis 3), as there was no difference in procedural learning based on the timing of the principle use. Unfortunately, the observed power was not high enough (0.42 where 0.80 is generally considered the minimum acceptable power to detect an effect when present) to determine that the non-significant effect was due to an absence of an effect. I therefore did not find any support for the idea that studying principles in the context of problem solving would be beneficial for procedural learning (measured with troubleshooting performance).

The secondary purpose of the first experiment was to compare two different methods of studying the principles: summarizing and reading. Prior studies on the effect of principles typically have not controlled how the principles are studied and this raises the question of what exactly participants are doing when they are provided with principles to study. Some might use self-explanation to process the principles whereas others might simply scan the information without any effort invested in learning the material. To evaluate the impact of providing principles in instructions it is necessary to consider the effect of the study method the learners use. If the principles are not sufficiently studied, they cannot be expected to increase learning.

If the principles are useful for learning (whether measured with tests of knowledge of the system or troubleshooting performance), it is important to guarantee that the information they provide is processed adequately. This is what I attempted to do.

The prediction was supported for knowledge of the system as participants summarizing the principles scored higher on the knowledge tests than the participants who read the principles (Hypothesis 2). The prediction was not supported for procedural learning as there was no difference between study method groups on the measure of troubleshooting performance (Hypothesis 4; again the observed power was insufficient to determine the source of the non-significant effect). The only benefit of summarizing the principles for task performance was that participants who had summarized the principles after completing the training tasks reported less subjective workload during testing than participants who had read the principles.

An interesting outcome of the first experiment was that when time needed to study the principles was taken into account and instructional efficiency was considered (relationship between the time needed to study and learning outcome), summarizing was less efficient than reading. Summarizing the principles resulted in higher scores on the knowledge tests than reading them, but the difference was only a few questions while the study time doubled. Therefore, there was a trade-off between time used to study the principles and knowledge test score and instructional designers should define which aspect of learning performance they want to emphasize in a particular context before implementing a study method.

The results of the first experiment speak to the relationship between knowledge of the system and procedural learning, and how principles affect these differently. Summarizing the principles led to an advantage on the tests of system knowledge, but this advantage did not translate directly into better procedural learning. One reason for this could be that the difference in knowledge of the system was not substantial enough to result in detectable differences in procedural learning. That is, summarizing the principles provided some additional knowledge (or better articulation of knowledge) over reading them or using them during task completion, but did not make a difference when completing tasks. However, this is not to say that principles are irrelevant for procedural learning. I did find that as the participants spent more time studying the principles their knowledge of the system increased, and better knowledge of the system was associated with better procedural learning. It seems therefore that studying principles matters for procedural learning, but maybe not as much as predicted by the mental model account, and requiring learners to summarize all the information might be unwarranted.

An interesting finding, or absence of finding, is the failure to find a difference between the Summarize-Before and Summarize-After groups on knowledge of the system and procedural learning. This means that the principles did not act as advance organizers when studied before training. It is difficult to provide an interpretation of a non-finding, but it appears that there were no obvious learning advantages of providing learners with the principles before starting the training tasks compared to after completing them. This suggests that the order of information presentation is not as important as often is supposed in the instructional literature. However, the issue needs further exploration.

There was no difference in training performance based on condition. Therefore, my hypothesis that using the principles during training task completion would lead to the best training performance (because these participants would have access to the information when needed in problem solving) was not supported (Hypothesis 5). These results are quite interesting, especially because the Summarize-After group was included to provide a baseline for training performance without principles. Therefore, it did not provide any advantage for training performance to have the principles. Unfortunately, as before, the observed power was not high enough (0.34) to allow me to conclude that the non-significant effect was due to absence of effect.

The same surprising results were found for ratings of subjective workload during training as there were no differences found among any of the groups. Again, the participants in the Summarize-After only had access to general procedural instructions and were therefore expected to report the highest workload of all the groups when completing the training tasks. These results raise the question of whether there is cost, in terms of workload, of integrating multiple sources of information that might offset advantages of having more information.

Another interesting finding concerning the Summarize-After condition is that the participants in this group spent more time studying the procedural instructions during training than the Summarize-Before and Use-During groups. This suggests that other sources of information affect how much the procedural instructions are studied. If only given procedural instructions, the learners will spend more time studying them than they would have if they had already spent time studying the principles (Summarize-Before) or if they had concurrent access to the principles (Use-During).

Because the first experiment was designed to rather rigidly control how each group used the principles, the design also had the problem of creating a rather artificial situation (e.g., participants in some groups not having access to principles at all during training) that is unlikely to be found in an authentic learning situation. It raises the question whether learners would spontaneously use the principles given differential constraints of the learning situation. The second experiment was designed to tackle this question.

EXPERIMENT 2

Introduction

A central hypothesis addressed in the first experiment was that principles are more helpful for procedural learning if used while problem solving during training than if they are studied separately. The second experiment was designed to define some factors determining whether people use the principles spontaneously for learning during training.

Prior research indicates that providing learners with general procedural instructions encourages them to use the principles to work out what they need to do, whereas having detailed procedural instructions will not. When provided with detailed procedural instructions, learners will follow them step-by-step without finding any need to consult the principles or meaningfully engage in the task (Catrambone, 1990, 1995; Duff & Barnard, 1990). The goal of the second experiment was to investigate whether the specificity of procedural instructions would influence how learners used the principles when completing tasks, in particular, whether giving participants general procedural instructions would encourage them to use the principles compared to giving them detailed procedural instructions. For the second experiment I also explored how the timing of principle use would influence learning in this context and therefore I proposed to use two of the conditions used in Experiment 1.

I had originally proposed to use the two conditions resulting in the best learning outcomes (declarative learning or knowledge of the system and procedural learning) in the first experiment. However, I found a host of problems associated with using this method when I tried to select the conditions to use for the second experiment. First, defining the “best” learning outcome was problematic because I used different dependent variables (e.g., for procedural learning) and in some cases different definitions of learning outcomes. For example, on knowledge of the system tests, the two summarizing conditions had a higher absolute score, but performance was better for the Read-Before and Use-During conditions when study time had been taken into account. Therefore, it was difficult to determine the “best” learning outcome in this case. Second, deciding what to do when there was not a significant difference among the conditions was difficult, as was the case with measures of procedural learning. I briefly considered using the averages to decide the issue (i.e., choosing the best based on averages regardless of whether differences were significant), but decided this would be problematic since the difference could be due to chance.

After unsuccessfully attempting to use “best” learning outcome to decide which conditions to use, I decided to base the selection on theoretical considerations. The design of the first experiment had essentially two main experimental groups based on theory: Summarize-Before and Use-During. There are theoretical reasons to believe that having organizing information beforehand is beneficial for learning (e.g., advance organizers) and this is what is traditionally expected – if not stated – in the literature on instructions. There are also theoretical reasons why using the principles during task completion might be beneficial (e.g., just-in-time information presentation, the task providing the context for knowledge construction). Therefore, I decided to select these two conditions to use in the second experiment to investigate how the timing of principle use and procedural instruction specificity influenced each other.

In the second experiment the manipulation of timing of principle use (before or during) was crossed with specificity of procedural instructions (general or detailed), creating four experimental groups. The *Summarize-Before-General* group summarized the principles before starting the training tasks and received general procedural instructions in training (this group was identical to the Summarize-Before group in the first experiment except for having access to principles during training). The *Summarize-Before-Detailed* group also had to summarize the principles before starting the training tasks, but received detailed procedural instructions in training. The *Use-During-General* group was not required to study the principles in any particular way, and could access the principles only during training along with general procedural instructions (this group was identical to the Use-During group in experiment one except for having access to the principles during training). The *Use-During-Detailed* group also only had access to the principles during training, but received detailed procedural instructions.

An important difference between the first and second experiments was that in the second all the participants had access to principles during training. I expected both manipulations – procedural instruction specificity and timing of principle use – to influence how the principles would be used during training task completion, and as a consequence, affect procedural learning.

I expected the participants in the two groups using the detailed procedural instructions (Use-During-Detailed and Summarize-Before-Detailed) to rely primarily on the procedural instructions during training and for the most part ignore the principles because the detailed procedural instructions provided all the information needed to complete the training tasks (leading to good training task performance). They would therefore not see any need to study this information, as their goal of completing tasks could be reached without it.

I expected the reliance on detailed procedural instructions to result in poor procedural learning for the groups using detailed procedural learning because the participants would neither study the principles nor think about what they were doing when following the detailed procedural instructions and completing the training tasks. I expected reports of low subjective workload would reflect this lack of effortful cognitive processing during training.

I expected the participants in the two groups using the general procedural instructions (Use-During-General and Summarize-Before-General) to engage in problem solving while completing the training tasks, leading to better procedural learning and less subjective workload during testing compared to participants using detailed procedural instructions.

However, I also expected that the participants in the Use-During-General group would demonstrate better procedural learning than participants in the Summarize-Before-General group because the Use-During-General participants would use the principles to help them in the context of problem solving, which the Summarize-Before-General participants were not expected to do.

Research has shown that people prefer to rely on memory rather than search out information, especially if searching out the information entails any effort, for example, by having to find the relevant piece of information and memorize it (Fu & Gray, 2006; Gray, Sims, Fu, & Schoelles, 2006). I therefore expected the Summarize-Before-General participants to be reluctant to study the principles during training, even if they had access to them, as they had already spent time summarizing them. This means that I had the same expectations for the comparison of these two groups (Summarize-Before-General and Use-During-General) in the second experiment as I had in the first, even if all the participants in the second experiment had access to the principles during training.

I expected both the timing of principle use and procedural instruction specificity to influence knowledge of the system: The participants in the Summarize-Before groups would, by summarizing, learn the information better than participants who were not required to study the principles in any particular manner (Use-During groups). The participants using general procedural instructions would be more likely to explore the system and engage with the learning materials in the process of solving the tasks, and therefore acquire better knowledge of the system compared to participants using detailed procedural instructions.

Based on these expectations I proposed five hypotheses about the learning outcomes in the second experiment.

Hypothesis 1: Participants in the Use-During-General group were expected to use the principles more than the participants in the other three groups during training, and I expected this to be reflected in the overall time used to study the principles during training task completion.

Hypothesis 2: Participants in the two groups receiving detailed procedural instructions were expected to generally show better performance on the training tasks than the participants who received general procedural instructions.

Hypothesis 3: I expected the timing of principle use (summarizing them before or using during training task completion) and procedural instruction specificity (having detailed or general procedural instructions) would both influence knowledge of the system. I expected participants who summarized the principles before training would show better knowledge of the system than those who did not, and the participants who received general procedural instructions would show better knowledge of the system than those who received detailed procedural instructions.

Hypothesis 4: I expected an interaction between timing of principle use and procedural instruction specificity for measures of procedural learning. I expected the participants using general procedural instructions would have better procedural learning outcomes than the participants using detailed procedural instructions. I also expected that the participants using general procedural instructions and only using the principles during training task completion (Use-During-General) would show better procedural learning compared to participants summarizing the principles before training (Summarize-Before-General).

Hypothesis 5: Ratings of cognitive load and task difficulty were expected to depend on whether general or detailed procedural instructions were provided. That is, the participants using general procedural instructions (Summarize-Before-General and Use-During-General) were expected to report higher cognitive load and task difficulty in training compared to the participants using detailed procedural instructions (Summarize-Before-Detailed and Use-During-Detailed). However, I expected the opposite pattern of results in testing: Participants using general procedural instructions would report less cognitive load and task difficulty than participants using detailed procedural instructions.

Method

The second experiment used the same simulation, tasks, and instructional setup as the first experiment. The procedure was the same as well, except that all the participants also had access to the principles during training.

Participants

A total of 98 undergraduate students at the Georgia Institute of Technology were recruited for the experiment. Their age range was be between 18 and 34 years (*M* = 20.5, *SD* = 2.6) and 48 were male (49%).

Students participating in the first experiment were not eligible to sign up for the second experiment, and students who had taken the course ECE 2040 at Georgia Tech, which covers circuit analysis, were not be eligible to participate either.

Participants were randomly assigned to an experimental condition (24 per condition) and compensated with course credit. During data collection it became clear that the data from two participants could not be used because of non-compliance. Therefore, these participants were replaced (hence a total of 98 participants recruited). Therefore, at the end of data collection the data from 24 participants in each condition was used, for a total of 96 participants.

Design

The experiment used a between-subjects design, where each participant experienced a single level of each of the two independent variables (timing of principle use and procedural instruction specificity). Timing of principle use refers to when the principles were used in the course of training: the participants were either asked to summarize the principles before starting the training tasks or only allowed to use the principles while completing the training tasks. Note however that all participants had access to the principles during training task completion, including the ones who summarized them beforehand (note: the participants did not have later access to the summaries they had created). Procedural instruction specificity refers to how specific descriptions were provided in the procedural instructions for the training tasks (general or detailed).

The two independent variables were completely crossed for a total of four different groups. The *Summarize-Before-General* group actively studied the principles before doing the training tasks by summarizing the main ideas and then received general procedural instructions during training. The *Summarize-Before-Detailed* group also studied the principles before doing the training tasks, but received detailed procedural instructions during training. The *Use-During-General* group was provided with the principles and general procedural instructions while completing the training tasks, and the *Use-During-Detailed* group was provided with the principles and detailed procedural instructions while completing the training tasks. The participants in the last two groups were not required to study the principles in any particular way.

The dependent variables for task performance were the same as in the first experiment: I measured the preferred learning modality for each participant and collected demographic information along with GPA and SAT scores at the start of the experiment. During training the instructional use was measured as the amount of time spent studying the instructions (both procedural and principles).

Task performance (both in training and testing) was measured as the time needed to complete tasks, the number of safety errors made, the number of unnecessarily replaced components, and the number of meter readings used. These procedural performance measures were standardized and then averaged to create a composite standardized measure of task performance. In addition, the hints needed to complete each task were counted and participants rated the subjective workload and difficulty of each task, and described what was wrong with the circuit after completing each task.

Declarative learning was measured with scores on two knowledge tests and at the end of the experiment I measured each participant's mental model of the circuit by asking them to draw the circuit from memory and to explain its functions.

The second experiment had three sessions: A training session where the participants completed the training tasks with the aid of instructions and experienced the experimental manipulations, and two testing sessions, one immediately following training (immediate) and one a week later (delayed).

Materials

The same materials used in Experiment 1 were used in Experiment 2: the simulation, tasks, instructions, hints, knowledge test, measures of subjective workload (abbreviated NASA-TLX) and task difficulty ratings, learning preferences questionnaire (VARK), questions on explaining and drawing the circuit.

The only difference between the experiments concerned the procedural instructions. In the first experiment all the participants used general procedural instructions, but in the second experiment half used general and half used detailed procedural instructions.

The detailed procedural instructions provided all the information of how to complete the task, including sub-steps, interim conclusions, and the exact actions required for each step. The general procedural instructions omitted detailed information, such as sub-steps and interim conclusions, and provided more details in the earlier than later steps. The creation of the detailed and general procedural instructions is described in Appendix A.

Procedure

The procedure for the second experiment was identical to the first experiment except that in the training session all the participants had access to the principles.

Preparing for Data Analysis

Data Removal and Outliers

Out of the 96 participants (24 in each condition) the data from one participant were removed during data analysis because he or she had not completed on the majority of the testing tasks (this was discovered after data collection had been completed). This left 23 participants in the Use-During-General condition, and 24 in all the other conditions. One participant in the Summarize-Before-Detailed condition did not return for the second session, which left 23 participants in the delayed session for that condition.

As in the first experiment, the difference in sample size was small enough to not affect precision of the estimation of the population mean, but a close attention was paid to Levene’s test of equal variance in the analyses to guarantee that the difference in sample size was not problematic.

Outliers were removed based on time-on-task. As in the first experiment, the upper outlier value was defined as being three standard deviations from the mean (about 50 minutes) and the lower value was defined as 10 seconds (based on expert performance). A total of 11 outliers (1.2%), or task instances, were removed.

Coding

As in Experiment 1, the following measures were coded: the summaries created by the participants, the open-ended knowledge test questions, and the drawings and explanations of the circuit. The coding schemes developed for the first experiment were used and the same raters coded the information in the second experiment. Therefore, the coding rubric descriptions and interrater reliability information provided for Experiment 1 also applies to Experiment 2 (see Appendix E).

Prior Knowledge

Prior domain knowledge was assessed in Experiment 2 in the same way as in Experiment 1 using the following measures: self-reported prior experience (whether participants had taken classes or done work related to circuitry), self-reported domain knowledge (participants rated their domain knowledge as none, some, or high), and a domain knowledge test (consisting of five topical questions). These were then coded in the same manner as in Experiment 1, creating three indicators of prior domain knowledge.

About half of the participants (47%) reported no experience (had not taken classes covering electrical circuits nor done any work on circuits), about 43% reported some experience (had either taken a class in high-school or college or done some work on circuits), and 10% reported having experience (taken one class or more and done some work). No difference in prior experience was found among the conditions ($p > .05$).

When rating their own knowledge of electrical circuits, 43% of participants said they had no knowledge of electrical circuits, 56% said they had some knowledge, and 1% said that they had good knowledge. There was no difference among the conditions in terms of self-reported domain knowledge ($p > .05$).

The scores on the five prior knowledge questions were coded as being either incorrect (0) or correct (1), by a single rater, and then aggregated into a score indicating domain knowledge. About third (34%) of the participants scored low, 40% scored in the middle range (2-3), and 26% scored high (4-5). When the participants in the four conditions were compared on the prior knowledge test score there was no difference among the groups ($p > .05$).

As these measures would be used in a regression and likely measure a similar construct, multicollinearity among the three prior knowledge measures was assessed (by regressing all the variables on each other). A R^2 of .8 or more was adopted as a criterion for an unacceptable level of multicollinearity (Lewis-Beck, 1980). The categorical variables were dummy coded and there were therefore altogether five variables representing prior knowledge (two representing the three levels of prior experience, two representing the three levels of self-rated knowledge, and one representing the score on the prior knowledge test). None of the R^2 values were found to be problematic as all were below .8 (.29, .37, .33, .17, .31) and therefore the three measures could be used together in multiple regressions to assess the effects of prior knowledge on the dependent variables.

Order Effects

Knowledge Test Order

As in Experiment 1, the participants completed a knowledge test on each of the testing occasions: one immediately after training and one a week later. The two versions of the knowledge test (version X and Y) were counterbalanced such that half of the participants received version X in the immediate session and half did so in the delayed session.

A t-test was used to determine whether the order of the knowledge test had an effect on the outcome on the test: whether administering version X in the first session and version Y in the second session (X-first) or vice versa (Y-first) would change the results. No order effects were found (X-first: $M = 20.69$, $SD = 4.34$; Y-first: $M = 21.62$, $SD = 4.47$; $p > .05$ in both cases) and the order of the knowledge test was therefore ignored in the analysis of the results.

Task Order

Half of the participants in each condition were trained with task set A and half were trained with task set B. The tasks sets were designed to be comparable and which task set was used in training was not expected to influence the results. This was assessed by comparing the learning outcomes of participants who used task set A in training with those who used task set B. There was no difference in knowledge test scores or averaged standardized measure of procedural learning between participants trained on task set A and task set B ($p > .05$). Therefore, which task set participants used in training was not taken into account in later analyses.

Each participant was randomly assigned to one of eight orders (four for each task set) which determined the order of the testing tasks. The order of the testing tasks did not have an effect on knowledge of the system or procedural learning outcomes ($p > .05$) and the order of testing tasks was not taken into account in the analysis of the results.

Results

The analyses were focused on the five hypotheses described earlier: instruction use during training (Hypothesis 1), procedural performance during training (Hypothesis 2), knowledge of the system (Hypothesis 3), procedural learning (Hypothesis 4), and ratings of subjective workload and task difficulty (Hypothesis 5). Additionally, auxiliary analyses were performed to investigate the: 1) learning outcomes in the immediate and delayed testing sessions, 2) quality of summaries made by the participants in the two summarizing conditions, 3) influence of preferred learning modality and demographic variables, and 4) drawings and explanations of the circuit.

As in Experiment 1, all analyses concerning whether participants knew what had been wrong with circuit after each task were moved into Appendix F, and all analysis concerning task difficulty ratings were moved into Appendix G. In both cases these measures provided little information: Most participants knew what had been wrong with the circuit and rated the tasks as difficult, and there were no differences among the groups on either one ($p > .05$).

In the second experiment, there were no differences based on the experimental manipulations of whether participants relied on memory to help them complete the retention tasks in testing (these were tasks they had already completed in training; $p > .05$). However, as in Experiment 1, the participants who relied on memory generally performed better on the retention tasks. As these results do not speak directly to the hypotheses tested in the experiment, I moved them to Appendix H.

Instructional Use (Hypothesis 1)

The time spent viewing instructions was recorded, both for the principles and the procedural instructions. The procedural instructions were only available during task completion, but the principles were available both during task completion and before training for the summarizing groups. Therefore, for the following analyses, principle use is considered both overall (all the time spent studying the principles) and during training task completion only.

In the first hypothesis I predicted that participants in the Use-During-General group would use the principles more often while completing the training tasks than the other three groups. I did however not make any predictions about differences in the time used to study the procedural instructions. Figure 13 shows the average time (in minutes) used to study principles and procedural instructions during training for each group. The time spent summarizing the principles before training is shown separately for the Summarize-Before groups.

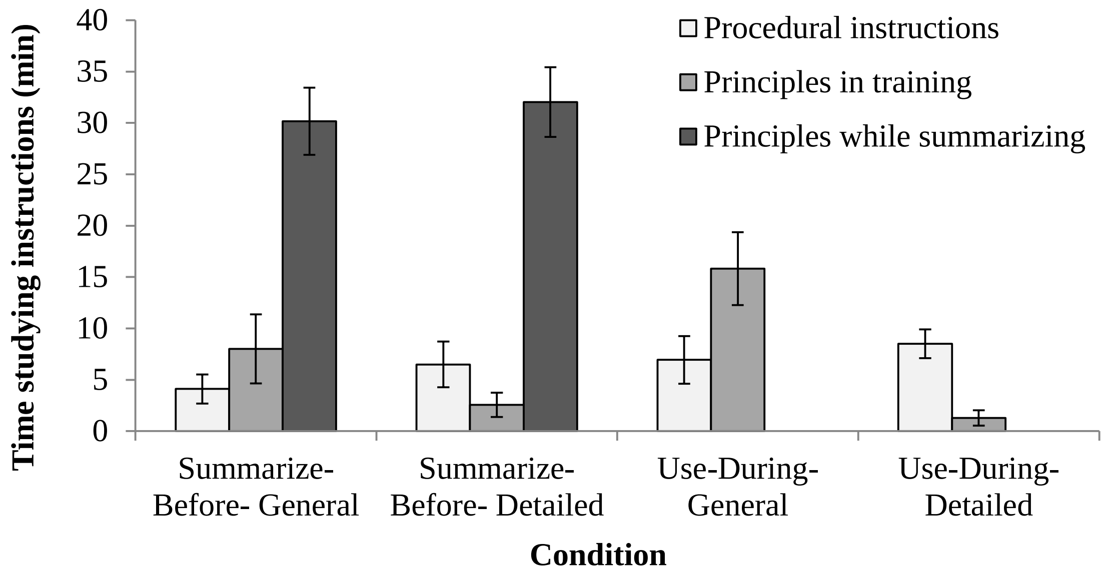


Figure 13. Time spent studying principles and procedural instructions in minutes for each condition. For the Summarize-Before groups, the time spent studying the principle is shown separately for summarizing and training task completion. The error bars show twice the standard error.

<u>Principle use</u>

The time spent summarizing the principles before starting the training tasks did not differ for the Summarize-Before-General and Summarize-Before-Detailed groups ($p > .05$), which is to be expected considering that there had been no treatment differences between the groups at the time the principles were summarized.

To assess the first hypothesis, I compared the Use-During-General group to the other three groups with a planned comparison, and found the hypothesis supported. The participants in the Use-During-General group used the principles more than the other groups during training ($F(1,659) = 4.27$, $p < .05$, $\eta_p^2 = .01$).

However, an omnibus analysis with the two independent variables revealed a significant interaction between the timing of principle use and procedural instruction specificity for the time studying the principles during training, $F(9,91) = 13.03$, $p < .001$, $\eta_p^2 = .13$. Based on the pattern of results shown in Figure 13, I followed up the interaction by comparing the simple effects of procedural instruction specificity separately for each level of timing of principle use.

For the participants in the Summarize-Before groups, the ones who used general procedural instructions spent significantly more time studying the principles in training than the ones using detailed procedural instructions, $F(1,46) = 9.32$, $p < .05$, $\eta_p^2 = .17$. The same results were found for the participants in the Use-During groups: The participants who had general procedural instructions spent significantly more time studying the principles in training than the ones who had detailed procedural instructions, $F(1,45) = 66.76$, $p < .001$, $\eta_p^2 = .60$. The interaction is therefore bound in the effect size as the difference between having general and detailed procedural instructions was considerably larger for the Use-During comparison than the Summarize-Before comparison. Therefore, having general procedural instructions led participants to consult the principles more during training, but the effect was more pronounced when the principles were only used during training task completion.

The hypothesis was supported: Participants using general procedural instructions and who only studied the principles during training task completion (Use-During-General) used the principles more while doing the training tasks than the other groups. However, the results were more complex than hypothesized as the difference between using general and detailed procedural instructions on the time used to study the principles during was considerably larger for the Use-During participants than the Summarize-Before participants.

Because of the difference found among the conditions, I investigated whether the time spent studying the principles during training would predict procedural learning. A multiple regression showed that the total time studying the principles did not predict procedural learning (standardized performance on the testing tasks), $R^2 = .001$, $p > .05$. The regression equation did not change when the time used to study principles only during training was used as a predictor.

<u>Procedural Instructions Use</u>

There was no interaction between timing of principle use and procedural instruction specificity for the time spent studying the procedural instructions ($p > .05$). There was, however, a main effect of both variables. The participants who summarized the principles before completing the training tasks (Summarize-Before groups) spent less time ($M = 5.30$ minutes, $SD = 4.67$) studying the procedural instructions than the participants who did not (Use-During groups; $M = 7.73$ minutes, $SD = 4.60$), $F(1,91) = 6.63$, $p < .05$, $\eta_p^2 = .07$. The participants who used general procedural instructions spent less time studying the procedural instructions ($M = 5.49$ minutes, $SD = 4.77$) compared to participants who used the detailed procedural instructions, ($M = 7.49$ minutes, $SD = 4.61$), $F(1,91) = 4.39$, $p < .05$, $\eta_p^2 = .05$. Because of the difference found for the time spent studying the procedural instructions, I investigated whether the time spent studying the procedural instructions predicted procedural learning using a multiple regression. The procedural instruction study time significantly predicted procedural performance on the testing tasks, $R^2 = .01$, $F(1,658) = 3.92$, $p < .05$, $\beta = 0.08$. As the participants spent more time studying the procedural instructions, the worse they performed on the testing tasks.

However, because there were two different versions of the procedural instructions, I repeated the analysis, but separately for the participants using general and detailed procedural instructions. For the participants using the general procedural instructions, the time used to study the procedural instructions did not predict procedural learning (R^2 = 001, $p > .05$), but it did for the participants using the detailed procedural instructions ($R^2 = .01$, $F(1,329) = 4.66$, $p < .05$, $\beta = 0.12$). That is, longer time studying procedural instructions predicted worse procedural learning, but only for the participants using detailed instruction.

Conclusions on Instructional Use (Hypothesis 1)

I had predicted that participants who used general procedural instructions and who only studied the principles while doing the training tasks would use the principles the most during training task completion. This hypothesis was supported; the Use-During-General group consulted the principles significantly more during training task completion than the other groups.

However, the time spent studying the principles during training task completion depended on both the timing of principle use and on procedural instruction specificity. The participants who were provided with the general procedural instructions always used the principles more when completing the training tasks compared to participants provided with the detailed procedural instructions. This difference was considerably larger for the Use-During groups than the Summarize-Before groups. Summarizing the principles before training tempered the effect of procedural instruction specificity on principle use during training. If the participants who used the general procedural instructions summarized the principles before the training, they used the principles less during training.

Procedural instruction use depended on both timing of principle use and procedural instruction specificity. The participants who studied the principles before starting the training tasks spent less time studying the procedural instructions than those who did not, and participants using general procedural instructions also spent less time studying the procedural instructions than those who used detailed. This reflects the difference in information reliance among the groups. People who summarized the principles beforehand probably relied more on their acquired knowledge when completing the task, and participants with detailed procedural instructions relied almost exclusively on them when completing the training tasks.

The overall time spent studying the principles did not predict procedural learning, but the time studying the procedural instructions predicted performance on the testing tasks, with those studying the procedural instructions more doing worse on the testing tasks. Upon further investigation, I found that it was the time studying the detailed procedural instructions that predicted performance on the testing tasks, not the time studying the general procedural instructions. Therefore, the participants who studied the detailed procedural instructions demonstrated poorer testing task performance. This is most likely due to the nature of the procedural instructions: Studying general procedural instructions is more helpful for procedural learning than studying detailed procedural instructions.

Training Performance (Hypothesis 2)

In the second hypothesis I predicted that the participants using the detailed procedural instructions would show better performance on the training tasks compared to participants using the general procedural instructions. Training performance was measured in the same way as performance on the testing tasks, with measures comprising time-on-task, number of safety errors, number of unnecessary components replaced, and number of meter readings. These measures were standardized and then averaged to create a composite task performance measure.

Generally, the participants using detailed procedural instructions showed better training performance than the participants using the general procedural performance (see Figure 14, and this difference was significant $F(1,281) = 42.46$, $p < .001$, $\eta_p^2 = .13$. The hypothesis was therefore supported. There was no main effect of timing of principle use ($p < .05$).

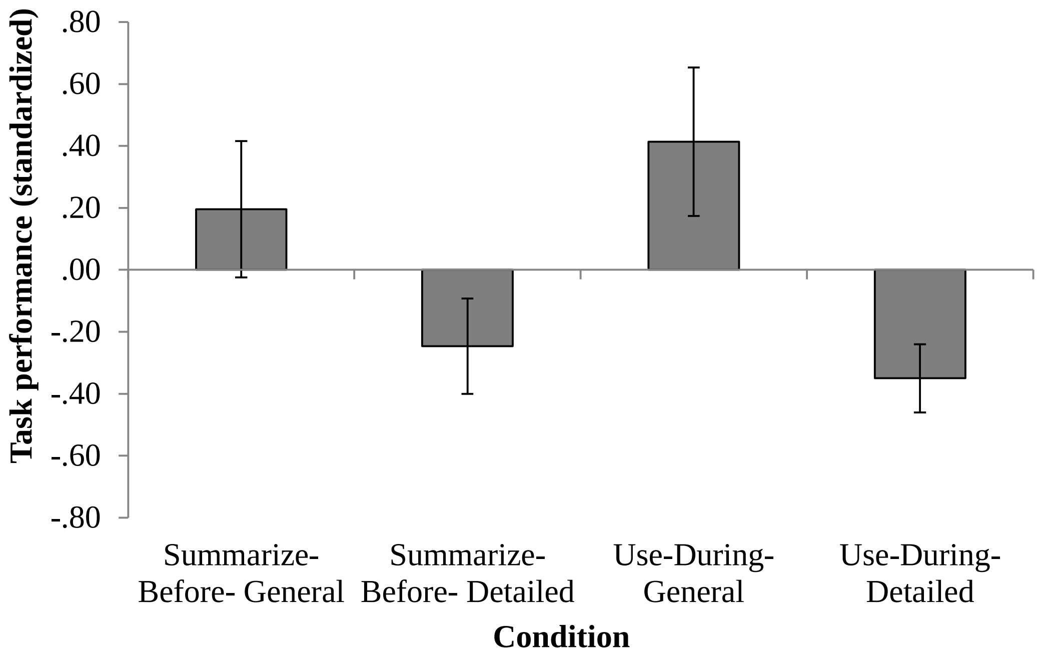


Figure 14. Average standardized task performance on the three training tasks for each condition. The error bars show twice the standard error.

A hierarchical multiple regression was used to determine whether procedural instruction specificity continued to have an effect on training task performance after prior knowledge had been accounted for. Prior knowledge variables were entered into the regression equation in the first step and condition in the second step (categorical variables were dummy coded). The regression showed that prior knowledge did not predict training performance ($R^2 = .01$, $p > .05$), but procedural instruction specificity did when prior knowledge had been accounted for ($R^2 = .14$, $\Delta R^2 = .13$, $F(6,269) = 7.43$, $p < .001$).

I had predicted that the participants using the detailed procedural instructions would show better performance on the training tasks than participants using general, and this hypothesis was supported. Prior knowledge did not account for any of the variance in training task performance, but procedural instruction specificity did when prior knowledge had been taken into account.

Knowledge of the System (Hypothesis 3)

Knowledge of the system was measured on the immediate and delayed testing occasions with parallel tests. Both tests had 18 multiple-choice questions and 18 open ended questions, for a total of 36 questions. Each question was scored as either correct (1 point) or incorrect (0 points) and then the number of correct answers was summed, so that on each test a participant could score from 0 to 36 points.

As in Experiment 1, no difference was found for knowledge test outcome on the two testing occasions (immediate session: $M = 21.87$, $SD = 4.68$; delayed session: $M = 21.74$, $SD = 4.43$; $p > .05$). Therefore, the results from the two testing occasions were collapsed for the analysis on knowledge test outcome and every participant was associated with two knowledge test scores in the following analyses.

In the third hypothesis I predicted that participants summarizing the principles before training would have higher scores on the knowledge tests compared to those who did not. In addition, participants using general procedural instructions were expected to score higher than those using detailed procedural instructions. That is, I expected a main effect of each independent variable.

There was however a significant interaction between timing of principle use and procedural instructions specificity, $F(1, 170) = 4.51$, $p < .05$, $\eta_p^2 = .03$ (see Figure 15). To unpack the interaction I analyzed the effects of procedural instruction specificity on the outcome on the knowledge tests separately for the Summarize-Before and Use-During conditions.

When the participants studied the principles before training, whether they used general or detailed procedural instructions during training did not influence knowledge of the system, $p > .05$. But, when the participants only had access to the principles during training (Use-During groups), the participants who used general procedural instructions scored significantly higher on the knowledge tests than those who used detailed ($F(1, 87) = 11.58$, $p < .001$, $\eta_p^2 = .12$).

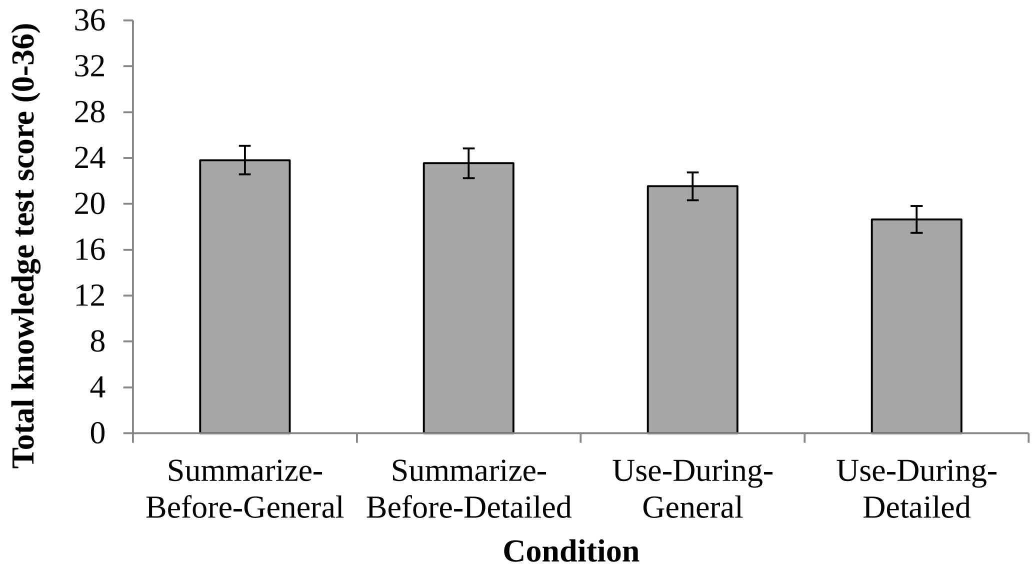


Figure 15. Average knowledge test scores (on the scale from 0 to 36) for each of the four conditions. The error bars show twice the standard error.

The hypothesis was not exactly supported as the results were more complex than hypothesized. I predicted a main effect of both manipulations, but the effect of procedural instruction specificity depended on the timing of principle use. If the participants summarized the principles before training, the procedural instruction specificity did not have an effect on knowledge test outcome. But, if the participants only had access to the principles during training, using general procedural instructions resulted in better outcome on the knowledge tests compared to using detailed.

Does the Principle Study Time Affect Knowledge Test Outcome?

Because of the differences found in time used to study the principles, I wanted to investigate the instructional efficiency of the groups by looking at whether longer time studying the principles delivered a proportionally higher score on the knowledge tests. To do this, I divided the total time taken to study the principles with the number of correct questions for each participant, thereby calculating the average principle study time per correct answer.

The participants in the Summarize-Before groups spent considerably more time studying the principles for every correct question on the knowledge tests than the participants in the Use-During groups (see Figure 16).

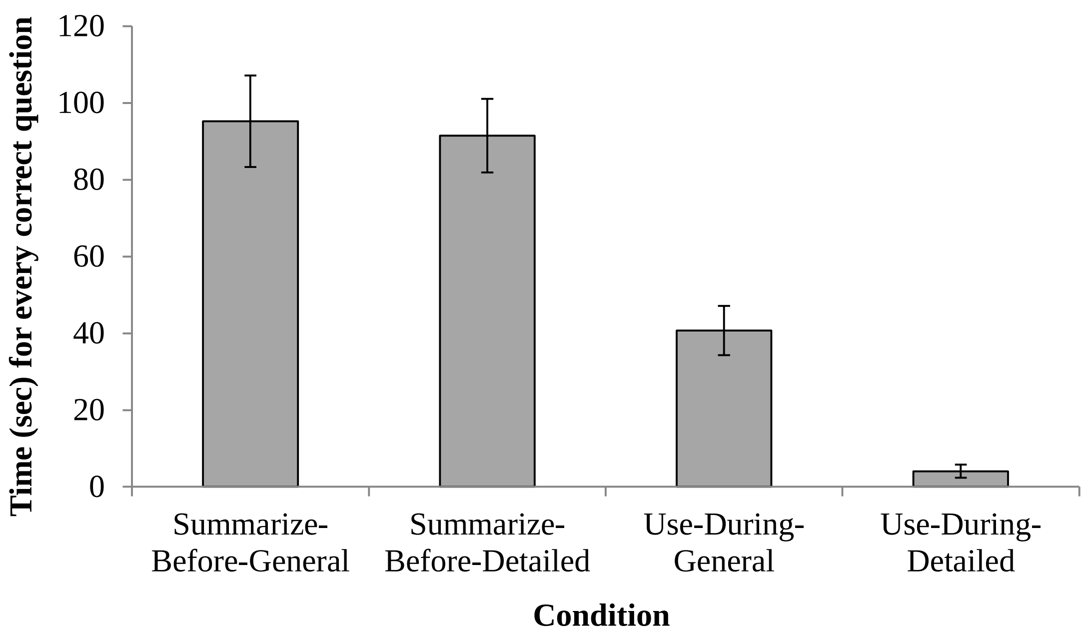


Figure 16. Time (in seconds) spent studying the principles overall, both before and during training, for every correct question on the knowledge tests. The error bars show twice the standard error.

There was however a significant interaction between the timing of principle use (Summarize-Before vs. Use-During) and procedural instruction specificity (General vs. Detailed), $F(1, 171) = 16.67$, $p < .001$, $\eta_p^2 = .09$. To investigate the interaction I calculated the simple effects for the Summarize-Before groups on one hand and the Use-During groups on the other.

There was not a significant effect of procedural instruction specificity for the Summarize-Before groups ($p > .05$), but there was a significant effect for the Use-During groups ($F(1, 88) = 136.76$, $p < .001$, $\eta_p^2 = .61$). The participants who had general procedural instructions (Use-During-General) spent significantly longer time studying the principles per each correct answer on the knowledge test compared to the participants who had detailed procedural instructions (Use-During-Detailed).

A hierarchical multiple regression was used to investigate to what degree condition affected knowledge test outcome over and above the time spent studying the principles. Study time was entered in the first step of the regression and condition (dummy coded with Use-During-Detailed as the reference group) was entered in the second step.

Time spent studying principles was a significant predictor of knowledge test outcome ($R^2 = .17$, $F(1,172) = 35.68$, $p < .001$), but when the time spent studying the principles had been accounted for, condition uniquely predicted knowledge test outcome ($R^2 = .22$, $\Delta R^2 = .05$, $F(4,169) = 11.74$, $p < .001$). This indicates that the difference among the conditions is not only due to the difference in the time spent studying the principles.

Does Prior Knowledge Predict Knowledge Test Outcome?

The participant's prior knowledge could be expected to influence the outcome on the knowledge test, and therefore I investigated whether condition would predict knowledge test outcome when prior knowledge had been accounted for.

Prior knowledge was measured with questions on prior experience, self-reported domain knowledge, and a test of prior knowledge. The prior knowledge variables were entered as predictors in the first step of a hierarchical multiple regression and condition was entered in the second step (the categorical variables were dummy coded). The regression showed that prior knowledge significantly predicted knowledge test outcome ($R^2 = .21$, $F(5,164) = 9.08$, $p < .001$), but condition also significantly predicted knowledge test outcome when prior knowledge had been statistically accounted for ($R^2 = .42$, $\Delta R^2 = .21$, $F(8,161) = 14.66$, $p < .001$).

Conclusions on Knowledge of the System (Hypotheses 3)

In the third hypothesis I predicted the participants who summarized the principles before starting the training tasks would score higher on the knowledge test compared to participants who used the principles only during training task completion. I also predicted that those who used general procedural instructions would score higher than those who used detailed procedural instructions.

The hypothesis was not quite supported, as there was an interaction between the timing of principle use and procedural instruction specificity. Whether participants used the general or detailed procedural instructions influenced the outcome on the knowledge test only when they just used the principle during training task completion, but procedural instruction specificity did not matter when the participants summarized the principles before training. That is, for the Use-During groups, having general procedural instructions led to a better outcome on the knowledge tests than having detailed procedural instructions.

Therefore, the specificity of the procedural instructions was important only when the participants used the principles when completing tasks, but not when they had already studied the principles. An explanation could be that participants in the Use-During condition had to engage in some problem solving when provided with general procedural instructions (using the principles for assistance) and this increased knowledge of the system. But when provided with all the details of how to complete the tasks the participants did not have to cognitively engage with the tasks. However, if participants had already studied the principles, what they did during training task completion mattered less.

This interpretation is supported by the findings that the participants in the Use-During-Detailed group spent less time studying the principles per each correct answer compared to those in the Use-During-General group (4 sec compared to 40 seconds). In addition, the participants summarizing the principles spent much more time studying the principles for every correct answer on the knowledge test compared to participants using the principles during training task completion (about 95 seconds compared to 40 and 4 seconds). These results are striking considering that the absolute difference among the groups in terms of score on the knowledge tests is at most five questions (out of 36).

The time used to study the principles predicted outcome on the knowledge tests, but so did condition even when study time had been accounted for. This indicates that the results are due not only to the time devoted to studying the principles. Prior knowledge also accounted for variance in knowledge test outcome, but condition did as well when prior knowledge had been statistically accounted for.

Procedural Learning Measured with Troubleshooting Performance (Hypothesis 4)

The fourth hypothesis predicted that procedural learning would depend on both the timing of principle study and the specificity of the procedural instructions, with those using general procedural instructions showing better procedural learning compared to those using detailed procedural instructions. Furthermore, of the participants receiving general procedural instructions, those who only had access to the principles during training task completion would show better procedural learning compared to participants who summarized the principles before training. That is, I expected an interaction between timing of principle use and procedural instruction specificity: The Use-During-General group was expected to show the best procedural learning, then the Summarize-Before-General group, and I expected the Use-During-Detailed and Summarize-Before-Detailed groups to show the poorest procedural learning (I did not predict a difference between these two).

Procedural learning was measured with troubleshooting performance immediately after training and a week later (difference between sessions will be addressed later), on both retention and transfer tasks. Performance was measured with time needed to complete tasks, number of safety errors, number of unnecessary errors, and number of meter readings. As before, these measures were standardized and then averaged into a composite standardized task performance score. In addition, whether participants needed hints to complete the tasks was assessed.

Overall the Use-During-Detailed group had the worst testing task performance, and the Summarize-Before-General group the best (see Figure 17).

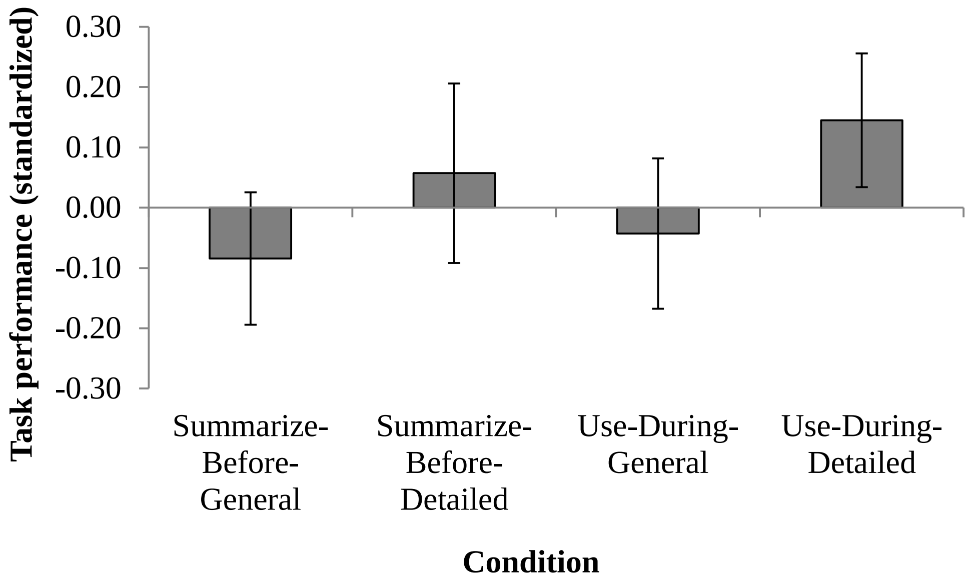


Figure 17. The average standardized testing task performance for each condition. The error bars show twice the standard error.

There was not a significant interaction between timing of principle use and procedural instruction specificity (p > .05) as had been predicted, but there was a main effect of procedural instruction specificity ($F(1, 656) = 7.02$, $p < .05$, $\eta_p^2 = .01$). The participants who used general procedural instructions performed better on the testing tasks (M = -0.06, SD = 0.75) compared to the participants who used the detailed procedural instructions (M = 0.10, SD = 0.84).

The part of the hypothesis that using general procedural instructions would be better for procedural learning than using detailed procedural instructions was supported: Having general procedural instructions led participants to perform better on the testing tasks. However, the part of the hypothesis stating that the effect of timing of principle use would depend on procedural instruction specificity was not supported. Timing of principle use did not affect procedural learning measured with performance on the testing tasks ($p > .05$).

Does Training Time Predict Testing Task Performance?

As there was a significant effect of procedural instruction specificity on testing task performance, I wanted to investigate whether the effects might have been caused by the time needed for completing the training tasks instead of the manipulation *per se*. A hierarchical multiple regression showed that training time did not significantly predict testing task performance ($R^2 = .003$, $p > .05$), but once training time had been statistically controlled for, the procedural instruction specificity manipulation did ($R^2 = .01$, $\Delta R^2 = .008$, $F(2,622) = 3.41$, $p < .05$).

Does Prior Knowledge Predict Testing Task Performance?

A multiple hierarchical regression, where prior knowledge variables (prior experience, self-reported domain knowledge, and outcome on the test of prior knowledge) were entered in the first step and the procedural instruction specificity manipulation was entered in the second step, showed that prior knowledge significantly accounted for time-on-task variance ($R^2 = .05$, $F(5,633) = 7.09$, $p < .001$), but when prior knowledge had been statistically accounted for condition uniquely did so as well ($R^2 = .07$, $\Delta R^2 = .02$, $F(6,632) = 7.68$, $p < .001$).

Type of Testing Task

Participants had worse performance on the transfer tasks ($M = 0.14$, $SD = 0.84$) compared to the retention tasks ($M = -0.14$, $SD = 0.72$; $F(1, 658) = 19.08$, $p < .001$, η_p^2 =

.03), but there was no interaction between the manipulations and test type (retention vs. transfer; $p > .05$).

Hints To Complete Testing Tasks

After working on the testing tasks for some time, the participants had access to three successive hints to help them complete the task. The final hint provided information about what exactly was wrong with the circuit and what component needed to be replaced to fix it.

Figure 18 shows the percentage of testing tasks where hints were used. The participants who used detailed procedural instruction during training were more likely to use hints to help them complete the testing tasks than participants who used general procedural instructions, and this was true for all three hints (Hint 1: $F(1, 657) = 14.63$, $p < .001$, $\eta_p^2 = .02$; Hint 2: $F(1, 657) = 14.17$, $p < .001$, $\eta_p^2 = .02$; Hint 3: $F(1, 657) = 11.83$, $p < .001$, $\eta_p^2 = .02$). There was neither a significant interaction nor a main effect of timing of principle use ($p > .05$ in both cases).

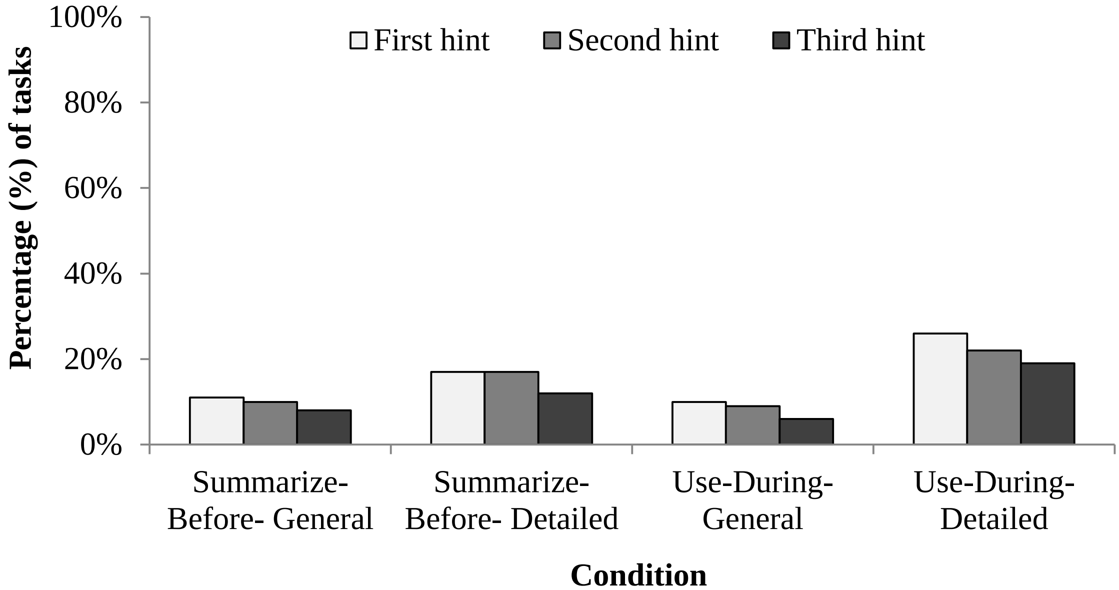


Figure 18. Percentage of participants in each condition who used the hints to complete the testing tasks.

Therefore, the hypothesis was partly supported: Using general procedural instructions made participants less likely to need hints to complete the testing tasks than using detailed procedural instructions.

Does Knowledge of the System Predict Procedural Learning?

The outcome on the knowledge tests depended on both the timing of principle use (Summarize-Before or Use-During) and procedural instruction specificity (General or Detailed) and I wanted to know to what degree this would translate into better procedural learning. To do so, I ran a multiple regression using the average score on the knowledge tests as a predictor for procedural learning.

The average score on the knowledge tests significantly predicted performance on the testing tasks (R^2 = .08, F(1,644) = 58.60, p < .001, β = -0.29): The higher the score on the knowledge test the better the performance on the testing tasks. The regression equation explained 8% of the variance in testing task performance.

Conclusions on Procedural Learning (Hypothesis 4)

I had predicted an interaction between the timing of principle use and procedural instruction specificity for measures of procedural learning: Using general procedural instructions would lead to better procedural learning outcomes than using detailed procedural instructions, but also that the Use-During-General group would show better procedural learning compared to the Summarize-Before-General group.

This hypothesis was only partly supported as general procedural instructions did lead to better procedural learning outcomes, but there was no interaction with or effect of timing of principle use. Therefore, using general procedural instructions seemed to encourage engagement with the tasks and problem solving during training, which subsequently increased procedural learning, when compared with using detailed procedural instructions specifying exactly how the training tasks should be carried out. These results were found even when training time and prior knowledge had been statistically accounted for. The same pattern of result was found for use of hints: The participants using the general procedural instructions were less likely to need the hints to complete the testing tasks than participants using the detailed procedural instructions. It has to be noted though that the effect sizes seen were small and this raises the question of practical significance of the results. But, given how little is generally known about which factors in instructions do determine procedural learning outcome; these results provide an important part of the puzzle.

Generally, participants showed worse performance on transfer tasks than retention tasks, and this effect was uniform across the conditions. The score on the knowledge tests predicted performance on the testing tasks, as the participants scored higher on the knowledge test the better the testing task performance.

Subjective Workload (Hypothesis 5)

I expected procedural instruction specificity to influence ratings of subjective workload and that these influences would be different for training and testing tasks. I expected the participants using general procedural instructions to report higher subjective workload than participants using detailed procedural instructions during training. But, that the opposite would be true for testing: Participants with detailed procedural instructions would report higher subjective workload than participants with general procedural instructions.

After each task, subjective workload was measured with an abbreviated NASA-TLX, which measures workload on six different dimensions. Five were used in the following analyses (mental demand, temporal demand, success at task, difficulty of obtaining that level of success, and degree of frustration). The sixth dimension, representing physical demand, was excluded. The rating of each dimension resulted in a score between 0 and 100 and these scores were averaged (the dimension for success was reversed first) to create a single general subjective workload score, where a lower number represented less subjective workload.

As in Experiment 1, the participants were given an example task to familiarize them with the simulation at the start of the experiment. Another purpose for having all the participants complete the same task before the manipulation took place was to get a baseline rating of subjective workload. The participants were explicitly told that this would be the easiest task in the experiment and to bear this in mind when rating workload. Subjective workload ratings were therefore not only considered in absolute terms, but also in relation to the baseline rating of the example task.

The overall range of average reported workload was 25 to 51 (out of 100; see Figure 19). The Use-During-General group reported the highest average workload during training and Use-During-Detailed during testing.

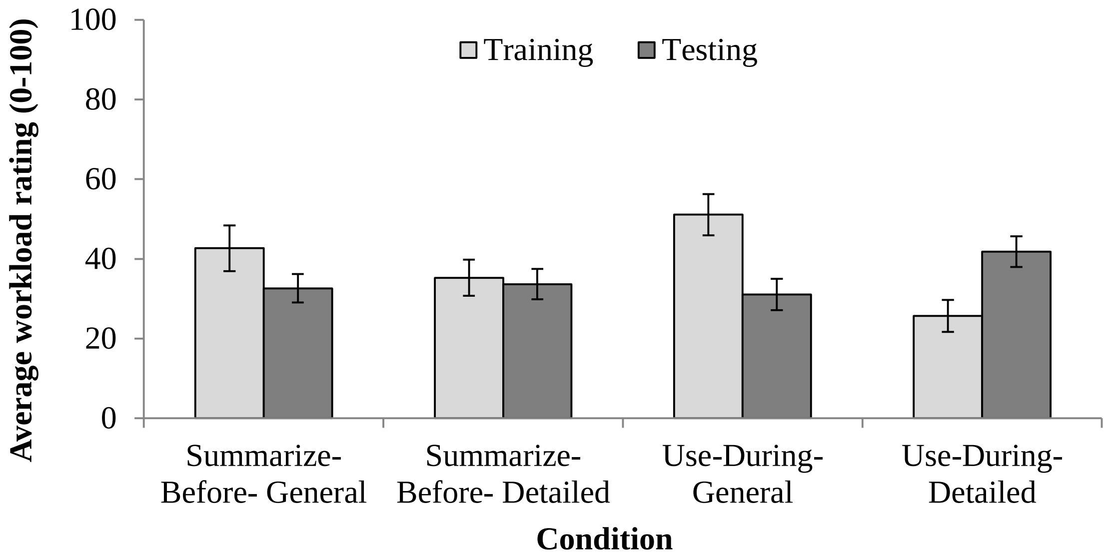


Figure 19. Average subjective workload ratings for training and testing tasks. The error bars show twice the standard error.

There was a significant interaction between timing of principle use and procedural instruction specificity for subjective workload ratings during training, $F(1,281) = 13.43$, $p < .001$, $\eta_p^2 = .05$. The interaction was followed up with simple effects analyses for each level of timing of principle use (Summarize-Before vs. Use-During). There was a significant difference between participants using general and detailed procedural instructions in both cases (Summarize-Before groups: $F(1,142) = 4.09$, $p < .05$, $\eta_p^2 = .03$; Use-During groups: $F(1,139) = 60.67$, $p < .001$, $\eta_p^2 = .30$), with participants using the general procedural instructions reporting higher subjective workload than participants using the detailed procedural instructions. The interaction was therefore based on the size of the differences: Procedural instruction specificity had a greater effect on subjective workload during training in the Use-During conditions than the Summarize-Before conditions.

There was also a significant interaction between timing of principle use and procedural instructions specificity for subjective workload reported during testing, $F(1,657) = 6.60$, $p < .05$, $\eta_p^2 = .01$. The interaction was again followed up with simple effect analyses based on the levels of timing of principle use (Summarize-Before vs. Use-During). There was no difference in reported workload in testing based on whether the participants in the Summarize-Before groups had used general or detailed procedural instructions ($p > .05$). But, for the participants in the Use-During conditions, the ones using general procedural instructions reported significantly lower subjective workload in testing than those using detailed procedural instructions ($F(1,327) = 15.56$, $p < .001$, $\eta_p^2 = .05$).

The hypothesis was generally supported: Participants who used general procedural instructions reported higher workload during training, but participants who used detailed procedural instructions reported higher workload during testing. The pattern of results was more complicated than expected, as this depended on the timing of principle use. In training, the difference between those using general and detailed procedural instructions was more pronounced for participants in the Use-During conditions compared to participants in the Summarize-Before conditions. In testing, the difference between those using general and detailed procedural instructions was only evident for participants in the Use-During conditions and not those in the Summarize-Before conditions.

When ratings of subjective workload were compared for training and testing, participants reported significantly higher workload during training ($M = 38.56$, $SD = 22.61$) than testing ($M = 34.85$, $SD = 24.54$; $t(944) = 2.19$, $p < .05$, $d = 0.14$, 95% CI [.38, 7.05]). In addition, when subjective workload ratings were compared for retention and transfer tasks in testing, the participants reported significantly higher workload when completing transfer tasks ($M = 38.20$, $SD = 25.91$) than retention tasks ($M = 30.37$, $SD = 21.82$; $t(659) = -4.21$, $p < .05$, $d = -0.33$, 95% CI [-11.47, -4.08]).

Subjective Workload with Baseline

The same method used in Experiment 1 was used to take baseline ratings into account in Experiment 2: The example task ratings were averaged into an average baseline rating for each participant and then this average baseline rating was subtracted from the average rating of each task for this participant. The resulting variable thereby represents the subjective workload over and above the baseline rating.

Figure 20 shows the average subjective workload ratings for the training and testing tasks when the baseline rating has been taken into account. As for the absolute ratings of subjective workload, the Use-During-General group reported the highest average workload during training and Use-During-Detailed during testing.

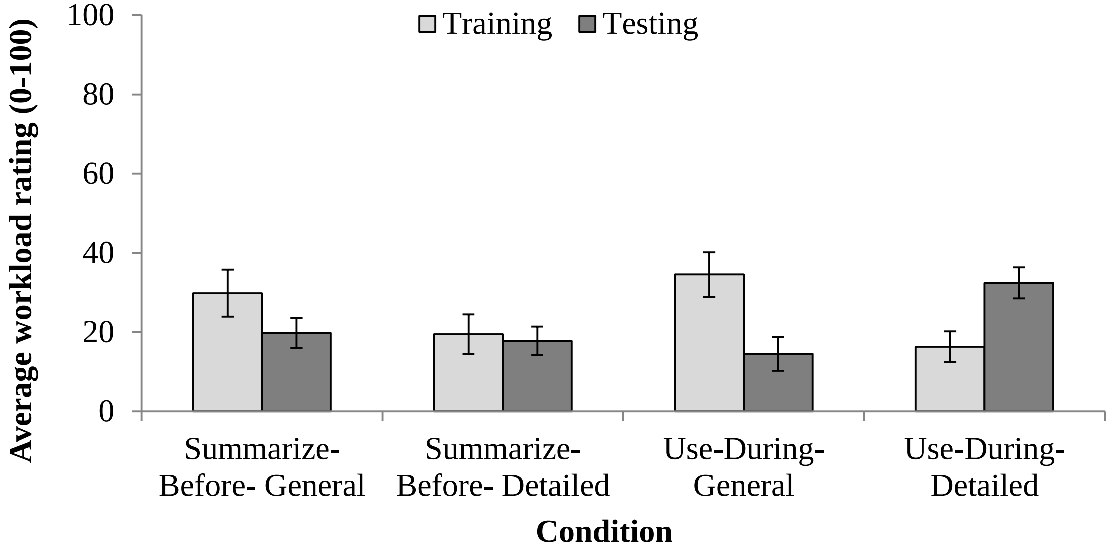


Figure 20. Average workload rating when baseline has been taken into account. The error bars show twice the standard error.

When rating the subjective workload of the training tasks, there was a main effect of procedural learning specificity, $F(1,281) = 30.77$, $p < .001$, $\eta_p^2 = .10$, and no effect of the timing of principle use (or an interaction; $p > .05$). The participants using detailed procedural instructions rated the subjective workload lower in training ($M = 17.90$, $SD = 18.95$) than participants using general procedural instructions ($M = 31.14$, $SD = 24.31$).

When rating the subjective workload of the testing tasks, there was a significant interaction between timing of principle use and procedural instruction specificity, $F(1,657) = 25.93$, $p < .001$, $\eta_p^2 = .04$. A follow-up with simple effect analyses for each level of timing of principle use, showed that participants in the Use-During groups using general procedural instructions reported lower subjective workload during testing than participants using detailed procedural instructions ($F(1,327) = 38.49$, $p < .001$, $\eta_p^2 = .11$), but there was no difference between those using general and detailed procedural instructions in the Summarize-Before groups ($p > .05$).

The hypothesis was therefore again mostly supported when baseline workload ratings had been taken into account. Participants using general procedural instructions reported higher workload in training than participants using detailed procedural instructions. In testing, the opposite was true for participants in the Use-During groups: Participants using general procedural instructions reported lower workload than participants using detailed procedural instructions.

Conclusions on Subjective Workload (Hypothesis 5)

In the fifth hypothesis I predicted that ratings of workload would depend on the procedural instruction specificity, with participants using general procedural instructions reporting higher subjective workload during training than participants using detailed procedural instructions, but the reverse would be true for testing.

This hypothesis was mostly supported: Participants who used general procedural instructions reported higher subjective workload during training, and this was true even when baseline ratings of workload where taken into account. The results for testing were a bit more complex than hypothesized as the effect of procedural instruction specificity depended on timing of principle use. The participants using general procedural instructions reported lower subjective workload in testing than participants using detailed procedural instructions, but this was true only for the Use-During groups (both in terms of absolute scores and when baseline had been accounted for). For the participants who had summarized the principles before starting the training there was no difference in reported workload during testing based on whether general or detailed procedural instructions had been used.

Therefore, for participants who only used the principles during training task completion, providing general procedural instructions added to experienced workload during training, but resulted in lower subjective workload when completing testing tasks (compared to participants who used detailed procedural instructions).

For participants who had summarized the principles before starting the training tasks, providing general procedural instructions added to workload during training, but there was no difference in reported workload based on procedural instruction specificity during testing. This means that providing general procedural instructions can have advantages in testing for learners who only use principles during training task completion, but it only adds to the workload of those who have already studied the principles before training and yielded no benefits during testing.

Participants rated overall subjective workload higher during training than testing and reported higher workload for the transfer tasks compared to the retention tasks, but in neither case did these results depend on the manipulations.

Changes over Time – Difference Between Testing Sessions

Participants were tested on two different occasions: immediately after training (immediate testing) and a week later (delayed testing). As in the first experiment, I did not predict an interaction between experimental manipulation and testing sessions, but I expected generally worse learning outcomes (both declarative and procedural) on the delayed testing session compared to the immediate one.

There was not a significant difference between the knowledge test scores on the immediate (M = 21.87, SD = 4.69) and the delayed (M = 21.74, SD = 4.43) testing occasions ($p > .05$). There was also not a significant difference between the procedural learning outcomes on the immediate (*M* = 0.05, *SD* = 0.81) and the delayed (*M* = -0.00, *SD* = 0.80) testing occasions ($p > .05$). The hypothesis was therefore not supported.

I had hypothesized that learning outcomes (both for knowledge of the system and procedural learning) would be worse on the delayed testing session compared to the immediate one, but this hypothesis was not supported. There was no difference found between testing occasions for knowledge of the system test scores or performance on the testing tasks.

Summaries Created By the Participants

The participants in the Summarize-Before groups summarized the principles before starting the training tasks. The summaries were then coded based on completeness (content points) and to what degree participants had rephrased the information (summary points). A higher number of content points meant the participants included more of the information from the principles in the summaries, and a higher number of summary points meant the participant had rephrased the information in their own words and were less likely to write the summaries verbatim.

The content and summary points were entered as predictors into a multiple regression and the knowledge test scores and the procedural learning measure were entered as the dependent variables. Summary points predicted outcome on procedural learning (see Table 11), with participants who summarized in their own words generally performing better on the testing tasks. However, the regression equation did not account for a significant amount of the testing performance variance. Content points did not predict procedural learning or knowledge test outcome.

Table 11
Summary of the multiple regression analysis predicting the knowledge test and procedural learning outcomes based on the content and quality of the summaries.

Predictor	Knowledge test score (β)	Testing task performance (β)
Content points	0.13	-0.01
Summary points	0.18	-0.12^{*}
R^2	.06	.01
F	2.56	2.73

$^{*}p < .05$. $^{**}p < .001$.

There was no difference between the Summarize-Before-General and Summarize-Before-Detailed groups in terms of content or summary points ($p > .05$).

The completeness and quality of the summaries made by the participants in the Summarize-Before conditions did not predict scores on the knowledge tests. But, participants who rephrased the principles when writing the summaries (the summaries were not verbatim) performed better on the testing tasks. This suggests that the benefit of summarizing information depends more on the learners putting the to-be-learned information in their own words than on including all the content.

Individual Differences

At the start of the experiment, participants were asked to provide demographic information (gender and age), their GPA and SAT scores, and they also completed the VARK questionnaire assessing their preferred learning modality. The effects of these variables on learning outcomes (knowledge of the system and procedural learning) was assessed with a multiple regression where the predictors were age, gender (dummy coded), GPA, SAT scores (verbal and math), and the VARK scores (four scores representing preferences for visual, aural, read/write, and kinesthetic modality).

Table 12 summarizes the regression results. There were three significant predictors for knowledge test score: gender, SAT-Verbal score, and the read/write score on the VARK questionnaire. Males were more likely to score higher than females on the knowledge test score and those who scored higher on the SAT-Verbal test and the read/write questions on the VARK questionnaire were more likely to score higher on the knowledge test.

Procedural learning was predicted by gender and the SAT-Verbal score. Males and participants with higher SAT-Verbal scores were more likely to show better performance on the testing tasks.

Table 12
Summary of the multiple regression analysis with individual difference variables predicting the knowledge of the system and procedural learning.

Predictor	Knowledge test score (β)	Testing task performance (β)
Gender	0.19*	-0.17*
Age	0.17	-0.05
GPA	0.04	-0.02
SAT-Verbal	0.19*	-0.10*
SAT-Math	0.02	-0.05
Visual score	-0.16	0.06
Aural score	0.12	-0.04
Read/write score	0.25*	-0.04
Kinesthetic score	0.12	-0.01
R^2	.20*	.05*
F	3.22*	2.97*

*$p < .05$. **$p < .001$.

Better knowledge tests scores and testing task performance were associated with being male and higher SAT-Verbal scores and higher knowledge test scores were associated with a preference for learning with textual based materials. The effect of verbal proficiency and preference for textual based learning is not surprising given the nature of the knowledge tests, but was less expected for the procedural learning outcome. One possible explanation is that because the instructions (both principles and procedural) were provided in a textual format, having good verbal or textual skills would give participants an advantage.

These results are different from what was found in Experiment 1, where preference for kinesthetic modality was associated with higher score on the knowledge tests and a preference for visual (non-verbal) modality was associated with a lower score. The results that a those with a preference for a non-verbal visual modality suffered on the knowledge tests and those with preference for learning from verbal information benefited, can be seen as two sides of the same coin. The principles were provided in verbal form and the knowledge tests were explicitly constructed from the principles. Therefore, those with preference for learning from verbal material would benefit and those with preference for non-verbal visual materials would suffer. Another reason for the different results could be the difference in the samples of the two experiments. That is, the participants in the first experiment had a greater preference for kinesthetic modality than the participants in the second experiment.

Drawings and Explanations of the Circuit

At the end of the experiment, participants were asked to draw the circuit and answer three questions requiring explanations of how the circuit worked. These were included to assess the participant's mental model of the circuit. The drawings and the explanations were coded for both correctness and how much detail they provided.

<u>Drawings of the Circuit</u>

The drawings were coded based on whether they depicted the structure of the circuit, provided details explaining the function of the circuit, and depicted details in general. As in Experiment 1, the scores for the circuit drawings were generally high for all four conditions; especially the structure and functional component scores (see Table 13). There was no difference among the groups for any of the three drawing components ($p > .05$ in all three cases).

Table 13
Means and standard deviations (in parentheses) on each component of the drawings (structure, functional details, and drawing details) for every condition. The possible range of scores is provided in parentheses in each column heading.

Condition	Structure points (0-26)	Functional detail points (0-7)	Drawing detail points (0-6)
Summarize-Before-General	24.79 (1.08)	5.25 (1.50)	2.96 (0.82)
Summarize-Before-Detailed	24.09 (3.06)	4.41 (1.97)	2.78 (1.04)
Use-During-General	25.05 (0.93)	4.00 (1.88)	2.73 (0.86)
Use-During-Detailed	25.04 (1.13)	5.07 (1.67)	2.72 (1.10)

Explaining the Circuit

The participants completed three questions where they were asked to explain how the circuit worked. First, they were asked to provide a short explanation of all the components of the circuit, using a picture of the circuit for reference. Second, they were asked to describe the workings of the relay in detail, again using a picture of the relay for reference. Third, they were asked to show where the circuit would be energized when the lights are off by drawing on a picture of the circuit.

On the first question, the Summarize-Before groups scored significantly higher than the Use-During groups (see Table 14; $F(1,91) = 15.60$, $p < .001$, $\eta_p^2 = .15$). There was no interaction between the two manipulations, and no difference based on the procedural instruction specificity ($p > .05$). On the second question (explaining the relay in detail) the scores were very low, indicating that participants generally had difficulty understanding, and hence explaining, how the relay worked (see Table 14). Again, there was a significant difference based on the timing of principle use ($F(1,90) = 9.76$, $p < .05$, $\eta_p^2 = .10$), with participants in the Summarize-Before groups scoring higher than the participants in the Use-During groups. There was no effect of procedural instruction specificity, and no interaction (p > .05 in both cases). On the third question (depicting where the circuit is energized when the lights are off), there was a significant effect of procedural instruction specificity ($F(1,90) = 6.32$, $p < .05$, $\eta_p^2 = .07$) and not the timing of principle use (p > .05). Participants using general procedural instructions scored higher on average than participants using detailed procedural instructions (see Table 14).

Table 14
Means and standard deviations (in parentheses) on each explaining question for every condition. The possible range of scores is provided in parentheses in each column heading.

Condition	Question1: Explain function of components (0-27)	Question 2: Explain how the relay works in detail (0-22)	Question 3: Depict where circuit is energized (0-7)
Summarize-Before-General	7.08 (3.94)	1.96 (1.43)	3.20 (3.06)
Summarize-Before-Detailed	8.17 (3.85)	1.92 (1.82)	1.40 (1.86)
Use-During-General	5.26 (2.86)	1.43 (1.59)	2.85 (2.65)
Use-During-Detailed	4.96 (2.85)	0.58 (0.65)	2.02 (2.44)

Conclusions on the Drawings and Explanations of the Circuit

As in Experiment 1, there was no difference among the conditions on any of the circuit drawings component scores. The drawings therefore did not reveal any difference in the mental models of the circuit that the participants might have.

The participants who summarized the principles before starting the training tasks were better at explaining the function of the circuit components and how the relay worked in particular, than participants who only used the principles during training task completion. These results suggest that having to summarize the principles resulted in better knowledge of the circuit components, or at least a better ability to articulate their function, than when the principles were only used in the context of task completion.

Participants who used general procedural instructions did better than participants who used detailed procedural instructions when asked to draw where the circuit would be energized when the lights are off. This suggests that the participants who used the general procedural instructions explored the workings of the circuit in more detail and, in the process, might have created a more complete mental model of how the circuit works, compared to being told exactly what to do in each step of completing the training tasks.

The first two questions were based on the information provided in the principles to a large degree, as the principles explicitly explained how each component worked. The third question (indicating where the circuit was energized) was, however, not based on the principles or any information provided in the instructions, but required the participants to have studied the circuit during task completion. Therefore, the results indicate, that the participants who summarized the principles gained better knowledge of how the components worked, whereas the participants who used the general procedural instructions gained better understanding of the circuit from problem solving experience.

Discussion

The focus in Experiment 2 was to investigate how procedural instruction specificity would influence the use of the principles in training and subsequent declarative and procedural learning. Participants were given either detailed or general procedural instructions while completing the training tasks and having access to the principles. Half of the participants had already summarized the principles before starting the training tasks, but the other half had access to the principles only during training.

The results were expected to show that using detailed procedural instructions during training would lead the participants to largely ignore the principles (Hypothesis 1), show better performance on the training tasks (Hypothesis 2), and experience less subjective workload during training (Hypothesis 5) compared to participants using general procedural instructions. Each of these predictions was supported: As predicted, whether participants used general or detailed procedural instructions did significantly influence principles use, task performance, and subjective workload in training. However, the results were more complicated than expected as the timing of principle use played a role. Using general procedural instructions led the participants to consult the principles more in training and report higher subjective workload, but this was especially true if they only had access to the principles during training. That is, the effect was not as strong when participants had studied the principles before the training. This shows that summarizing the principles before starting training tempers the effect of using general procedural instructions. This tempering occurs most likely by influencing learners to rely on the knowledge of the system already acquired – due to the advance summarizing of the principles – and, as a result, they experience less subjective workload because they have less information to study and integrate while figuring out how to solve the tasks.

I predicted that using detailed procedural instructions would lead to worse knowledge of the system (Hypothesis 3), poorer test task performance (Hypothesis 4), and greater subjective workload (Hypothesis 5) at testing than using general procedural instructions. These predictions were made because participants using the detailed instructions would be less likely to study the principles and become cognitively engaged with the training tasks than participants using the general procedural instructions. I also expected that when using general procedural instructions, participants who summarized the principles would show better knowledge of the system and worse procedural learning than participants who had access to the principles only during training.

These predictions were partially supported: Using general procedural instructions did lead to a higher score on the knowledge tests and lower reported workload, but only when the principles were used just during training and not when they had been summarized before. Using general procedural instructions did also lead to better testing task performance and less use of hints at testing, but unlike what had been predicted, there was no influence of timing of principle use.

To summarize: Task performance, both during training and testing, was influenced by procedural instruction specificity. Participants using the detailed procedural instructions had better performance at training but worse at testing compared to the participants using the general procedural instructions. Knowledge of the system and subjective workload during testing were influenced by both timing of principle use and procedural instruction specificity. Whether the participants used general or detailed procedural instructions did not influence knowledge of the system or reported workload when the participants had summarized the principles before starting the training. When the participants had access to the principles only during training, being provided with general procedural instructions led both to better performance on the knowledge tests and less subjective workload at testing.

Therefore, the procedural instruction specificity influenced the use of principles, the knowledge of the system, task performance and subjective workload (both during training and testing). The use of principles – along with declarative and procedural learning – increased when the participants used the principles, at the cost of more workload and worse task performance at training. The timing of principle use also played a role. Asking the participants to summarize the principles before training tempered some of the negative effects of using detailed procedural instructions. When considering the use of principles, these results demonstrate the importance of paying attention to the other information provided in the instructions, particularly the level of detail in stepwise procedural instructions.

GENERAL DISCUSSION

The purpose of the two studies was to understand factors determining whether providing principles in instructions for procedural tasks is helpful for declarative learning (knowledge of the system) and procedural learning (performance on troubleshooting tasks without instructions).

The Mental Model Account

According to the mental model account, principles should be helpful for learning because learners can create a comprehensive model of how the system works by studying the principles, and establish good knowledge of the system (enhancing declarative learning). Furthermore, having good knowledge of the system is then expected to assist learners in figuring out what they need to do when faced with new tasks, as they can deduce what actions to take from how the system works, thereby increasing procedural learning (Bibby & Payne, 1993; Borgman, 1999; Duff & Barnard, 1990; Karreman & Steehouder, 2004; Kieras & Bovair, 1984; Mayer, 1981; Patrick & Haines, 1988). Figure 21a shows a schematic depiction of the traditional mental model account, where the solid arrows represent the direct causality assumed.

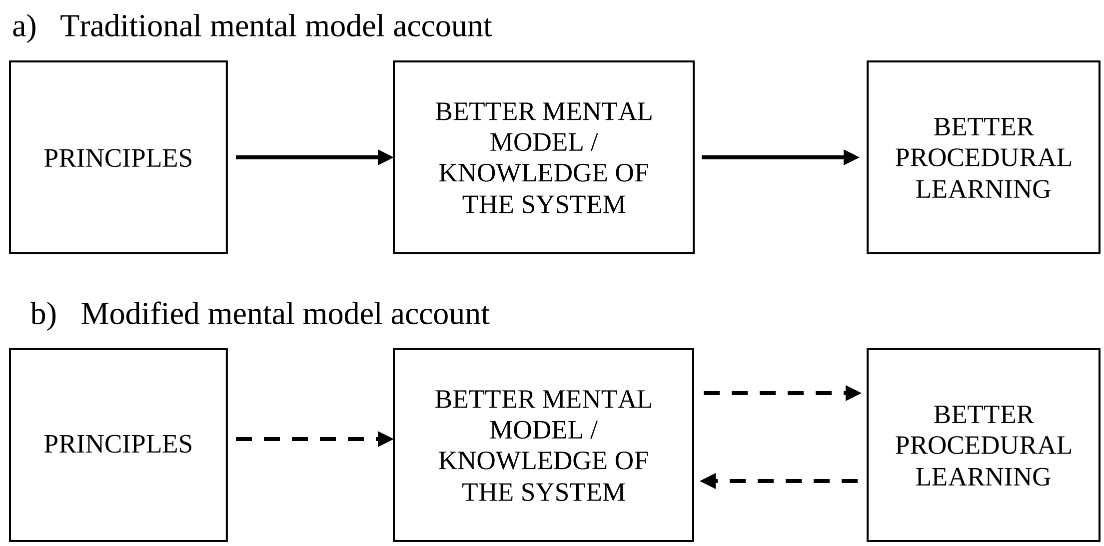


Figure 21. a) *Traditional mental model account*: Providing principles leads to better mental modal or knowledge of the system, and this in turn leads to better ability to complete tasks. The solid arrows between the boxes represent assumptions of directional causality. b) *Modified mental model account*: Having a better mental model or knowledge of the system can lead to better procedural learning, but better procedural learning can also lead to better mental model or knowledge of the system. The dotted arrows represent a qualified relationship.

Research on including principles in instructions has not yielded unequivocal results, suggesting the mental model account might be too simplistic and other factors are important in this context. The experiments were designed to explore the role of three factors that could be expected to influence whether providing principles is advantageous for learning: Timing of principle use (study before, during, or after doing tasks), the method used to study the principles (read or summarize), and the influence of procedural instruction specificity (whether the procedural instructions are general or specific).

The idea that providing principles in instructions for procedural tasks is beneficial because learners will acquire better knowledge of the system was supported in both experiments, but it depended on the study method. Participants who summarized the principles (either before or after completing the training tasks) scored higher on the knowledge tests than participants who either read the principles or used them during training task completion. Therefore, providing principles helps learners establish better knowledge of the system but only when the principles are actively studied, such as by summarizing (see Figure 21b for a schematic depiction of the modified mental model account where a dotted arrow – marked 1 – between providing principles and having a better mental model represents this qualified relationship).

It was interesting however that summarizing was not a particularly efficient study method, as the participants summarizing the principles took more time to study the principles per each correct answer on the knowledge tests than the other participants. This finding has two implications: First, teachers or instructional designers have to consider the trade-off between the time required to study and the test outcome when deciding on the method to use as this relationship is not necessarily linear. Second, the knowledge of the system acquired during task performance by using *procedural instructions* should not be under-estimated. Even the participants who spent very little time studying the principles (e.g., Use-During-Detailed group studied the principles for only 80 seconds on average) still managed to answer about half the questions on the knowledge tests correctly even though the questions on knowledge tests were based on the principles.

The mental model account includes the assumption that having better knowledge of the system translates into better performance on tasks within the system, especially tasks involving troubleshooting or problem solving. The present study does not provide strong evidence for this assumption: Even though the participants summarizing the principles showed better knowledge of the system in both experiments, they did not show better procedural performance on the testing tasks. However, the effect might be more subtle than the overall comparison among the conditions might reveal. In both experiments more time studying the principles was associated with better outcome on the knowledge tests, and better outcome on the knowledge test in turn predicted better performance on the testing tasks (though accounting for very little of the variance). In neither experiment did the time studying the principles directly predict better performance on the testing tasks. Therefore, if learners study principles effectively (as evidenced by better performance on some measure of knowledge of the system), they are more likely to show better performance when completing tasks without the aid of the instructions. It is not enough though to just provide the principles and expect performance on tasks to increase. In addition, there are various important factors still unaccounted for in this relationship, and further research is needed. Figure 21b shows a schematic depiction of the modified mental model account where a dotted arrow from having a better mental model to better procedural learning is used to represent this qualified relationship (arrow marked 2).

Advance Organizers

A prediction based on timing of principle use, and one implicit in the traditional mental model account, was that the principles could act as advance organizers for the learning taking place during training (e.g., Ausubel, 1960; Mayer, 2003). No evidence was found supporting this prediction. A comparison of the participants summarizing the principles before and after training in Experiment 1 did not show any advantage, in terms of knowledge of the system or procedural learning, of summarizing the principles before as opposed to after having completed the training tasks. If anything, slight advantages were found for summarizing the principles after training, as these participants were better able to explain how the relay worked compared to the participants summarizing the principles before training. Perhaps completing the training tasks provided the participants in the Summarize-After group a context for studying the principles or led them to pay attention to specific parts (i.e., the relay) that they had found complicated during training task completion. That is, perhaps the training tasks served as advance organizers for studying the principles and the experience of completing tasks can serve as organizing structure for later knowledge acquisition. This relationship is represented in Figure 21b by a dotted arrow leading from better procedural learning to having better mental model or knowledge of the system (arrow marked 3). There is therefore reason to believe the assumed causal directionality inherent in the traditional mental model account might not be a valid assumption.

Active Engagement

As has been mentioned, the participants who summarized the principles had better knowledge of the system than those who did not. This indicates that active engagement with the materials through some activity such as summarizing results in better learning than passively reading the information. The findings did provide an interesting insight into how summarizing works: It was the quality of the summaries, not the completeness of the content, that predicted performance on knowledge tests (Experiment 1) and procedural learning (Experiment 1 and 2). Rephrasing information therefore was more important than covering all the information in the principles. This suggests that merely telling learners to summarize information might not be as effective for learning as teaching them to put the information in their own words, because some learners will interpret summarizing to mean they should repeat the main ideas verbatim. This is similar to recent findings showing that rephrasing is a more effective method for study than taking verbatim notes (Bujak, 2010).

The idea that active engagement with the learning materials is important has also been advocated for procedural learning. For example, the minimalist approach to instructions emphasizes that learners should construct their own knowledge through experience with doing the tasks (e.g., Carroll, 1990) and others have advanced the idea that exerting effort and including some difficulty in a learning situation is important for learning and skill development. Robert Bjork has advocated the idea of desirable difficulty (Bjork, 1994, 1999; Schmidt & Bjork, 1992), where some level of difficulty during learning is desirable to foster cognitive processes necessary for effective learning (e.g., by spacing practice over time, using tests as learning events, interleaving learning sessions with other learning activities, varying the material to be learned, and decreasing feedback). On a similar note, in his work on deliberate practice, which is based on the practices of experts, Anders Ericsson has claimed that deliberate practice is inherently effortful (Ericsson, et al., 1993). The current results support these ideas. In the second experiment, the participants who used general procedural instructions struggled more with the training tasks than the participants who used detailed procedural instructions and reported higher subjective workload. In testing however, the opposite was the case: The participants who used general procedural instructions showed better performance on the testing tasks, and reported less workload even if they had not summarized the principles. These results suggest that to enhance procedural learning, the instructional materials should encourage learners to invest cognitive effort, such as through problem solving or increasing difficulty, and giving learners general procedural instructions is one way of doing that.

In Experiment 1 a surprising result was that performance on the training tasks was not affected at all by whether the participants received the principles. That is, there was no difference in training task performance between the Summarize-After group and the other groups. One explanation for these results could be found in procedural instruction specificity. The test task performance results in Experiment 2 showed an advantage of using general procedural instructions, and I have suggested that this is due to the problem solving that the participants had to engage in when completing the training tasks. Perhaps, this problem solving or active engagement in the tasks is the reason for the equal performance seen in Experiment 1, as all the participants used general procedural instructions.

Even if some problem solving was beneficial for procedural learning, using the principles in the course of this problem solving was not. There was no evidence that using the principles during training task completion had any advantage during training or testing. I had proposed that principles should be helpful when used during task completion, as the information would be provided exactly when needed by the learners (based on ideas of Just-In-Time information presentation) (Kester, et al., 2006; Kester, et al., 2001). But, no evidence was found that using the principles while completing tasks, as opposed to studying them outside of task completion, helped procedural learning.

Retention and Transfer

In both experiments participants showed worse performance and reported more subjective workload when completing transfer tasks compared to retention tasks. The participants therefore had more difficulty solving transfer tasks than retention tasks. There was no interaction between the manipulations and the type of testing task (retention vs. transfer) in either experiment. Therefore, I found no evidence that timing of principles, the method used for studying the principles, or procedural instruction specificity had differential effects on retention or transfer. The mental model account suggests principles should be especially helpful for transfer of learning because the learner can draw on his or her knowledge of the system to assist in problem solving, but the results here do not support this.

Differences between Testing Sessions

A very surprising result in both experiments was the lack of difference across sessions, both on knowledge of the system tests and performance on testing tasks. It is difficult to interpret the absence of effect (null results), especially because the observed power was low (0.28 in Experiment 1 and 0.11 in Experiment 2), but as these results were found for both experiments, I believe there might be a reason to suspect a genuine finding. It is difficult to explain these results, but the reasons might include the length of the retention interval, testing effect, and the role of the simulation as a retrieval cue. The retention interval used was a week, but in this case this might not have been long enough to produce the expected differences between the training and testing occasion, especially as the participants did complete a testing session (immediate testing) between training and the delayed testing. Recent research has shown that testing can by itself become an important learning event and even lead to better retention for learners than when they are given another study opportunity (Roediger & Karpicke, 2006a, 2006b). This testing effect has been effectively demonstrated using textual materials, but there is reason to believe that the testing effect could also be seen for procedural learning (Van Gog & Kester, 2011). Therefore, the immediate testing occasion might in and of itself have strengthened learning, resulting in the better outcome on the delayed testing session than expected. In addition, the simulation itself might have acted as a substantial retrieval cue for the participants, helping them remember what they needed to do when completing the tasks.

Conclusions

The question addressed in the experiments was what factors determine whether providing principles in instructions for procedural task helps learning? I examined three main variables: The timing of principle use, the method of studying the principles, and the role of procedural instruction specificity. Timing of principle use was found to have very little effect overall and the evidence does not support recommending any particular presentation timing of the principles. The method of study did influence knowledge of the system, and there was some evidence that having good knowledge of the system enhanced procedural learning as well (if less than the mental model account would lead one to expect). Therefore, making sure the learners actively study the principles is one factor that makes providing principles in instructions for procedural tasks helpful. The specificity of the procedural instructions was found to be quite important and determined whether learners engaged in problem solving during training task completion and whether they consulted the principles. Therefore, whether learners use the principles spontaneously and during problem solving is determined by the nature of the procedural instructions. General procedural instructions coupled with good principles can produce effective transfer at a relatively modest cost during training.

The findings of this research have implications in domains where people need to learn how to use a system that is governed by principles and where the principles can help them deduce what they need to do to control the system in new situations. The challenge for further research in this area is to investigate how different features of the system itself (e.g., distance between user controls and inner components, coherence and complexity of the system) and the nature of the tasks that people perform (e.g., task complexity or step dependence) interact with the implementation of principles and specificity of procedural instructions in training and test performance.

APPENDIX A

Creating the Instructions

All instructions for the tasks used in the experiment were created from hierarchical task analysis. The task analysis was used to identify the information and actions needed to complete each task, as well as commonalities and differences between tasks. This information was then used to create three types of instructions: Detailed procedural instructions, general procedural instructions, and principles.

The task analysis was used to identify goals and sub-goals of each task and the information and actions needed to be able to reach the goals and sub-goals. This information was used to created detailed procedural instructions that listed every step (goal) and sub-step (sub-goal) of carrying out the tasks. In addition, the detailed procedural instructions included the conclusion to be made from each step. For example, a step is to “check whether fuse is blown”, a sub-step is to “use voltmeter to test whether fuse is receiving voltage by measuring at FU-1”, and conclusion is “the voltmeter reading should show 115 V, indicating that the fuse is receiving voltage”. Detailed procedural instructions were created for all seven tasks used in the experiment from the task analysis information.

The general procedural instructions were created from the detailed procedural instructions with an algorithm. This helped guarantee that all the general procedural instructions were created in the same way. The algorithm included two main parts: First, for each step only the main step and final conclusion of the step were included (omitting all sub-steps and interim conclusions). Second, the earlier steps included more detailed information than later steps (see Table 15 for an example). The second part of the algorithm was found necessary after pilot testing various algorithms for creating the general procedural instructions from the detailed ones. I found that if the directives in the steps (exactly which components to test and replace) and the conclusions (what the testing revealed about the problem) were too detailed later in the instructions on a single task the participants did not engage in problem solving, but relied on the information without trying to understand how the conclusions were reached and next step decided. This was confirmed in debriefing when one pilot participant stated that when he did not understand exactly what was going on in a step in the general procedural instructions, he opted to move on to the next step instead of trying to figure it out. By fading out the amount of information provided in steps the learner is required to follow the logic of the method and engage in problem solving to figure out exactly where to test and how to fix the circuit, which is the goal of providing the general procedural instructions. Table 15 provides a comparison of general and detailed procedural instructions of the same task.

Table 15

An example of detailed and general procedural instructions for a single task (task O1).

Detailed procedural instructions	General procedural instructions
1. Try pressing the ON-buttons and note what happens ⇨ The lights don't turn on and the relay does not energize	1. Try pressing the ON buttons and note what happens ⇨ The lights don't turn on and the relay does not energize
2. Use the Voltmeter to check whether the fuse is blown a. Open the meter (using the "meter" button on the top panel) b. Turn the voltmeter on (turn the dial to the V symbol on the meter) c. Put the black lead on ground (the lower terminal of TB1-G; Note: the terminal is the screw) d. Put the red lead on the FU-2 terminal ⇨ The meter reading will be 115.0 V, indicating that the fuse is not blown ⇨ This tells you that the problem is most likely an open in the circuit and you should use the voltmeter to locate the problem	2. Use the Voltmeter to check whether the fuse is blown ⇨ The meter reading will tell you that the problem is most likely an open in the circuit and you should use the voltmeter to locate the problem
3. Divide circuit in two, (1) ON-and OFF-button area and (2) relay and light area, to locate the problem a. Using the Voltmeter, put the red lead on PB4-3 and try pressing the ON-buttons ⇨ The meter reading will show 0.0 when the ON-buttons are pressed, indicating that the problem is in the button area (because the relay is not getting any voltage)	3. Divide circuit in two, (1) ON-and OFF-button area and (2) relay and light area, to locate the problem ⇨ The meter reading will indicate that the problem is in the button area (because the relay is not getting any voltage)
4. Rule out the ON-button area a. Using the Voltmeter put the red lead on PB3-2 and press each ON-button ⇨ The meter reading should be normal: will show 0.0 volts by default and change to 115.0 volts when the ON-buttons are pressed	4. Divide the button area in two and rule out one half by testing at the midpoint

Table 15 continued

5. Locate the problem within the OFF-button area a. Put the red lead on PB5-3 and press any ON-button ⇨ The meter reading should be normal: will show 0.0 volts by default and change to 115.0 volts when the ON-buttons are pressed	5. Locate the problem within the identified area by dividing the area into two and testing at the midpoint ⇨ The meter reading should allow you to identify the component that is the probable cause for the fault
6. Test the probable cause: The wire between PB4 and PB5 a. Put the red lead on PB4-3 and press any ON-button ⇨ The meter reading shows that the problem is in the wire: will show 0.0 Volts when the ON-buttons are pressed	6. Test the probable cause
7. Turn off and lock out the current a. Turn the multimeter off by pressing the *OFF* position on the meter b. Zoom in on the breaker panel (click on the panel when the mouse is a magnifying glass) c. Turn off the current for the lighting circuit by clicking on the switch marked “Lighting circuit” d. Press the *Lock Out* button at the bottom of the panel e. Close or move the breaker panel window	7. Turn off and lock out the current
8. Refasten the wire between PB4 and PB5 a. Select the screwdriver from the top panel b. Use the screwdriver to unscrew the wire at the PB5-3 terminal ⇨ You will see a Loose Connection message at this point – this is the reason for the malfunction c. Select to tighten the connection …	8. Fix the fault by refastening the component …

The information from the task analysis was used to create the principles to make sure that the principles were relevant to the tasks that the participants had to complete. After creating the principles, the general procedural instructions were used to verify that the principles were indeed relevant to the tasks. That is, each step in each of the procedural instructions was connected to the relevant principles, and only the principles that were identified as being informative for a procedural step were included. Table 16 provides examples of some of the principles that were included. An effort was made to make sure the principles include only generalities and rules in the domain (on how to troubleshoot the circuit) and not tips for use or other types of declarative information.

Table 16
Example of principles provided in the experiment.

Principle	Description and excerpts
The relay	Describes how the relay works *Excerpt*: The relay coil receives voltage through R1-2 and this means that a reading at R1-2 will show whether the relay coil is receiving voltage.
The fuse	Describes the function of the fuse *Excerpt*: The fuse is a device that is designed to protect the circuit when too much current flows through it. A short in the circuit can create an excessive current, and this should cause the fuse to open (the fuse blows), interrupting the flow and isolating the circuit.
The voltmeter	Describes the purpose of the voltmeter, how to use it, and what voltmeter readings indicate. *Excerpt*: Voltmeter is the best tool to detect an open in the circuit because there should be voltage present above the location of the open and not below the location of the open
Electrical faults: Open and shorts	Describes the two types of faults, open and shorts, and how best to detect them (meter tool and readings). *Excerpt*: An open is when a broken wire, loose connection, burned out component creates a break in the current path of the circuit (prevents current from flowing). … A short occurs when two or more isolated components come into contact with a grounded object (e.g., insulating on wires gets bad) and the current travels on a different path from what was originally intended. This can cause an excessive flow of current in the circuit which would blow the fuse.
Divide and eliminate	Describes a method for locating the problem in the circuit, the most logical division of the circuit and the diagnostic testing points to eliminate parts from the identified problem area. *Excerpt*: When there are no symptoms to provide any hints about the location of the problem, a systematic approach to locating the problem is helpful. One approach is to use a “divide and eliminate” method, where you break the circuit into areas and conduct tests at the point that divides one area from another (dividing point) to eliminate parts of the circuit as problem locations

APPENDIX B

Task Descriptions

Task Sets

A total of seven different tasks were used in the experiment. Participants completed three in training and the four remaining were used for testing transfer. Two sets of tasks were created and counterbalanced across conditions: set A included tasks O1, O4, and S3 as training tasks, and set B included O3, O2, and S1 as training tasks. These tasks were chosen for the two sets because the tasks in each set match on the type of fault and method required for completion.

In tasks O1 and O3 the fault was an open due to a wire that is loose (same fault) but the faulty wires were in different places in the circuit. Tasks O4 and O2 also had an open in the circuit, but in both these tasks the problem was in the relay. In addition, O1 and O2 required the learner to understand how to section the circuit to locate the problem, whereas O4 and O3 required the learner to understand what a specific piece of diagnostic information meant (that the relay was energized in O4 and that the fuse did not receive current in O3).

In tasks S3 and S1 the fault was a short in the circuit, and in both cases the short was in a wire (but in different places in the two tasks). These tasks both taught the participants to diagnose a short in the circuit (starting with the fuse being blown) and use the ohmmeter to locate the problem.

The two sets of tasks therefore comprised of tasks that taught similar aspects of the task domain and this means that the set not used for training could be used as a transfer test of procedural learning for the first set.

Task Order

There were four different task orders for each set of training tasks covering every combination of testing tasks with some constraints creating a total of eight task orders in the experiment. The constraints were that in the immediate testing session the first task was a retention task where the fault was an open, the second was a transfer task where the fault was an open, and the third task was a transfer task where the fault was a short. In the delayed testing session the first task was a retention task where the fault was an open, the second was a retention task where the fault was a short, the third was a transfer task where the fault was an open, and the fourth was a transfer task where the fault was a short. The eight orders of tasks had every possible combination of task order within these constraints (see Table 17).

Table 17

Counterbalancing order of testing for each training tasks set. Set A consists of tasks O1, O4, and S3. Set B consists of tasks O2, O3, S1. R stands for retention tasks and T stands for transfer tasks.

		Immediate Testing			Delayed Testing			
Order	Training	Task 1 (R)	Task 2 (T)	Task 3 (T)	Task 1 (R)	Task 2 (T)	Task 3 (T)	Task 4 (T)
1	Set A	O1	O2	S1	O4	S3	O3	S2
2	Set A	O4	O3	S1	O1	S3	O2	S2
3	Set A	O1	O3	S2	O4	S3	O2	S1
4	Set A	O4	O2	S2	O1	S3	O3	S1
5	Set B	O2	O1	S3	O3	S1	O4	S2
6	Set B	O3	O4	S3	O2	S1	O1	S2
7	Set B	O2	O4	S2	O3	S1	O1	S3
8	Set B	O3	O1	S2	O2	S1	O4	S3

Participants in each experimental condition could therefore be assigned to one of the following task order: A-1, A-2, A-3, A-4, B-1, B-2, B-3, B-4, where A and B refer to the training task set and the numbers refer to the testing task order. This was used to counterbalance task order within each experimental condition and allowed me to determine whether there were any effects of the training tasks used and the order of testing tasks.

APPENDIX C

Development and Description of the Hint System

The hint system was implemented after pilot testing showed that some participants had difficulty completing the tasks, especially the testing tasks where they did not have any instructions to aid them. The goal of the hint system was to increase the likelihood that all participants completed each task for both practical and methodological reasons.

Practically, providing hints to allow the participants to complete tasks guaranteed that the simulation logged the dependent variables for the task (number of meter readings etc.) as these measures are not logged in the simulation unless the task is correctly completed. It also guaranteed that the time needed to complete tasks was kept contained. Pilot testing showed that if tasks had not been solved after the maximum par-value duration the participants were usually floundering and becoming increasingly frustrated. At that point the pilot participants usually started to ask the experimenter for help, and when queried, revealed that they had no idea how to solve the task or what to do next. It is doubtful that allowing more time on task under those circumstances would have increased the likelihood of them completing the task. Instead, it most likely would have resulted in the participants not completing the experiment in the allotted time. The experiment was already taking a long time (5 hours total) and increasing the time was undesirable for various reasons (e.g., participant fatigue). Therefore, standardizing a hint system would limit the time spent in unproductive frustration per task and provide an objective way of delivering help to the participants instead of having the experimenter field questions and provide help on a case by case basis.

Methodologically, the hint system offset to some degree the otherwise increasing difference in learning between participants who successfully complete the testing tasks and participants who were unsuccessful and could not complete the testing tasks (different testing effects). That is, participants presumably learned more from a testing task if they eventually figured out what the problem was, otherwise what they did and experienced while trying to complete the task was rather uninformative. This means that participants who successfully completed a testing task could relate the meter readings and symptoms to a specific and known problem whereas the participants who were unsuccessful at the task did not have that opportunity. With each successive testing task that is either successfully completed or not this difference in learning would in all probability have increased. By implementing a hint structure I wanted to decrease any learning difference due to differences in feedback during testing. This is not to say that there will not be inherent differences in amount of learning between participants that are more or less successful at completing the testing tasks, but it does level the playing field somewhat because all participants eventually got feedback on what the problem was.

In addition, the hint system allowed me to objectively define when participants failed to complete a task; if they required a hint to tell them exactly what was wrong with the circuit to fix it then they could be considered to have failed at completing the task on their own.

Initially, I had only intended to include the hints for the testing tasks, but in one case a pilot participant had considerable difficulty completing the third (and most difficult) training task and repeatedly asked the experimenter for help. Therefore, I found it prudent to develop a hint system for the training tasks as well for the reasons listed above.

The hint structure for the testing tasks included three hints for each task of increasing specificity and they appeared one at a time after a certain amount of time had elapsed.

The first hint did not appear until after the defined maximum par-value time had elapsed for each task. For example, for a testing task with 15 minute maximum par-value the hints would not start appearing until after 15 minutes had elapsed. The second hint appeared three minutes after the first, and the third appeared three minutes after that. This means that in general the testing tasks took at most about 21 minutes (opens) or 26 minutes (shorts) to complete if the participants decided to use the hints. They of course had the option of ignoring the hints and take longer to complete the tasks. As described before the hint structure therefore provided me with a metric for defining whether participants were unsuccessful at the task. If they needed the third task to be able to complete the task the task was considered unsuccessful.

Every task had three hints associated with it: the first hint told the participant whether the problem was an open or a short (and which meter to use to locate the problem), the second hint indicated the general location of the problem (e.g., supply area, ON-buttons, OFF-buttons, relay, or lights), and the third hint told the participant exactly what the problem was (e.g., replace wire between PB2-1 and PB3-1). Each hint appeared as a button on which the participant could click to see the hint. Therefore, the participants could decide whether they wanted to use the hint or not.

The hint system for the training task was implemented in the same way as the hint system for the testing task with one exception: The hints did not appear until after a considerably longer time. I believed that an increase in time was sensible as in training the participants had the instructions to help them and more time would be needed to read the instructions. In addition, having a hint system in place allowed the experimenter to tell participants who turned to him/her for help that hints would eventually appear and that they should try to use the information provided to complete the task. This was found necessary as some participants seemed to prefer to ask for help instead of using the instructions. For the training tasks the first hint did not appear until after 40 (open) or 45 (short) minutes had elapsed.

APPENDIX D

The Time Limit for Studying the Principles

The participants in the Read-Before groups controlled how much time they used to study the principles, but the Summarize-Before participants had up to 45 in total to read and summarize the principles. This upper limit was implemented after two pilot participants in the Summarize-Before condition took over an hour to summarize the principles (62 and 66 minutes respectively). An exit interview indicated that this was due to them being stressed about adequately learning the information in the principles as they knew they would not have access to them later. This posed somewhat of a problem as it extended the time needed to complete the experiment considerably.

To solve the problem, I decided to create an upper limit for each principle to prevent the participant for taking too much time summarizing the principles. At the same time I wanted to provide them with ample time to do so. In trying to reach these two goals at the same time I averaged the time the two pilot participants spent summarizing each principle and compared with the time a participant in the Summarize-After condition needed to summarize (going on the assumption that the participant in the Summarize-After condition would know what information was important based on the experience with doing the tasks). For each principle I used the halfway/median point between the average of the Summarize-Before participants and the time the Summarize-after participant needed. For example, the average of the Summarize-Before pilot participants for the principle on the relay was 10 minutes whereas the Summarize-After participant only needed 4 minutes to summarize it. I therefore decided on 7 minutes as the allowed maximum time for summarizing this principle. After deciding on the times for all eleven principles (summing to 45 minutes) I made sure this time had not been exceeded by the Read-Before participants I had already collected data from. Their overall time ranged between 11 and 22 minutes.

I implemented the maximum allowed time in the program by providing a countdown clock on the screen for each principle, showing the participants in the summarizing conditions how much time they had left for summarizing. They therefore could use the countdown to help pace themselves. Trying this on the next few participants in the Summarize-Before condition it turned out all of them completed summarizing within the time limit, did not consider the time too short, and the summaries they generated were neither verbatim nor hurried.

APPENDIX E

Coding

Coding was required for some of the measures used in the experiment: the summaries created by the participants in the Summarize-Before and Summarize-After conditions, the open ended knowledge test questions, and the drawings and explanations of the circuit.

Summaries

The summaries created by the participants in the Summarize-Before and Summarize-After conditions were based on whether they included the main ideas of each principle and whether they had been summarized in the participants own words or typed verbatim. The rationale behind the coding scheme was that a summary in the participant's own words including all the main ideas conveyed in the principles would indicate better learning than an incomplete summary typed verbatim. In the first case the participant would be required to work with the learning materials and get engaged in processing the information at a deeper level than they would in the second case (Chi, 2009; Pressley, 2006).

Each principle was coded for content and one point was given for every identified piece of information in the principle. For example, the principle on the relay had four content points; two concerned with the relay coil (one describing how the relay coil received voltage and one describing the return path from the relay) and two concerned with the seal-in contacts (one describing how they receive voltage and one describing the structure of the seal-in contacts). The principles could give two (type of errors) to eight content points (ohmmeter), for a maximum of 43 content points for all 11 principles.

Each principle was also coded for how it was summarized. If the participants did not summarize the principle (e.g., only typed the title) they would get 0 points (an error message would appear if they did not write anything and tried to navigate from the page). If the summary was mostly or completely verbatim they got 1 summary point, but if it was mostly or completely in their own words they would get 2 summary points. A participant could therefore get 0 to 22 points for how well they summarized the principles.

Two raters coded the summaries and to calculate interrater reliability they coded the same set of six summaries. An absolute agreement intraclass correlation coefficient (ICC) was used to assess interrater reliability because then the value of the scores is taken into account along with agreement on the rank ordering of scores (McGraw & Wong, 1996; Shrout & Fleiss, 1979). An interrater reliability criterion of .8 was adopted (Cohen, 1988; Landis & Koch, 1977).

The absolute agreement ICC for the two raters was .94 on the content points (95% CI: .18, .99; $p < .001$), and .40 for the summary points (95% CI: -.31, .88; $p > .05$). This shows good interrater reliability for the content of the summaries, but not for the summary points. The raters recalibrated the coding scheme for the summary points and again coded a set of six summaries (different from the first set used to calculate the ICC and different from the ones used for recalibration). This time the ICC for the summary points was .89 (95% CI: .48, .98; $p < .05$) and high enough to reach criterion.

Rubric for Coding Summaries

<u>Circuit = 5 points</u>

(1) Under normal operating conditions, pressing any of the ON-buttons (PB1, PB2, or PB3) energizes both the relay (R1) and the lights (LT1 and LT2) – turns the lights on. Likewise, pressing any of the OFF-buttons (PB4, PB5, or PB6) de-energizes the relay and lights – turns the lights off.

(2) The relay keeps the lights either energized or de-energized after the buttons are released.

(3) When the lights are off (under normal operating conditions), the circuit is only energized to the ON-buttons (PB1-1, PB2-1, and PB3-1) and the relay's seal in contacts (R1-1 and R1-8) which are always energized.

(4) The ON-buttons are by default open (create a break in the path of the current) and keep the circuit below them from receiving current. When ON-buttons are pressed they close and complete the circuit, energizing the relay. When the relay gets energized the two seal-in contacts in the relay close and even if the ON-button is released (becomes open again) the circuit stays energized because the seal-in contacts are closed. Once the two seal-in contacts close the lights are energized and turn on.

(5) The OFF-buttons are by default closed, meaning that they complete the circuit unless they are pressed. When an OFF-button is pressed it creates an open in the circuit and de-energizes the relay. When the relay is de-energized the two contacts will open, de-energizing the lights which turn off. When the OFF-button is released the relay stays de-energized because the seal-in contacts are now open. Once the seal-in contacts are open the lights are de-energized and turn off.

Fuse = 4 points

(1) The fuse is a device designed to protect the circuit when too much current flows through it. Typically a short to ground causes excessive current to flow and that in turn causes the fuse to open (the fuse blows). This interrupts the flow of the current, isolates the circuit and protects the components in the circuit that could otherwise be damaged. If you find a blown fuse chances are there is a short to ground in the circuit, and you should investigate the cause. The fuse can be in one of three possible states:

(2) The fuse is working: It receives full voltage (115.0 V) and delivers full voltage (115.0/114.8 V) to the rest of the circuit.

   ⇨ If the fuse is working when there is a fault in the circuit it means that there is an open somewhere in the circuit below the fuse.

(3) The fuse is blown: It receives full voltage (115.0 V) but delivers no voltage (0.0 V). That is, the fuse creates an open and the current does not run through the fuse.

   ⇨ If the fuse is blown it means that there is a short in the circuit.

(4) The fuse is not receiving voltage: In this case you will see no voltage (0.0 V) on either end of the fuse and this is why it is always necessary to measure whether the fuse is receiving voltage.

   ⇨ If the fuse is not receiving voltage it means that there is a problem in the circuit before the fuse.

Relay = 4 points

The relay (R1) consists of the relay coil and two seal-in contacts:

(1) The relay coil receives voltage through R1-2 and this means that a reading at R1-2 will show whether the relay coil is receiving voltage. You can also see whether the

relay coil gets energized by whether the dot in the middle of the relay fills up or not. It should fill up when energized (when ON-buttons pressed). Even if the relay coil is receiving current (and gets energized) the problem can still be in the relay's seal-in contacts

(2) The return path from the relay coil is R1-7. In the tasks you will do, it is never going to be informative to test at R1-7 because the voltmeter reading at R1-7 is always 0 V.

(3) The two seal-in contacts receive voltage at R1-1 and R1-8. These points are always energized if the fuse is working.

(4) The first seal-in contact is R1-3 and R1-4: if the contact is open (not letting current through the relay) R1-4 will be energized (but not R1-3), but if the contact is closed (letting current through the relay) R1-3 will be energized (and not R1-4). The two points of the seal-in contact are therefore never energized at the same time. The second seal-in contact is R1-5 and R1-6 and it works the same way as the first.

Faults: Open and shorts = 3 points

(1) Electrical faults are either: (1) components that have become open or (2) components that have become shorted to ground. The faults you encounter can be either an open or a short.

(2) An *open* is a break in the current path in the circuit (prevents current from flowing). It can be the result of a broken wire, loose connection, burned out component, etc. Note however that some components are designed to create an open – such as the fuse or the buttons. The voltmeter is the best tool to find an open in a circuit

(3) A *short* occurs when two or more isolated components come into contact with a grounded object (e.g., insulating on wires gets bad) and the current travels on a

different path from what was originally intended. When a short occurs, over-current devices such as a fuse will open the circuit to protect it, and this is why a blown fuse is often an indicator of a short. The ohmmeter is the best tool to find a short in a circuit.

Grading note: For points 2 and 3, only ½ point it they only mention the tool used to find the circuit and don't define open/short

<u>Symptoms = 4 points</u>

Before you do anything it is informative to test how the circuit operates, because sometimes what happens (the symptoms) can reveal diagnostic information. There are two components of the circuit that can change when you press the ON-buttons: the relay and the lights.

(1) If the *relay is energized* (the small dot in the center of the relay fills) you know that it is receiving current and the problem is in the relay or below (if the lights don't turn on).

(2) If the *lights turn on* at all you know that they are receiving current

If the symptoms are not informative (nothing at all happens when the ON-buttons are pressed), you should

(3) first check the fuse to see whether the problem is an open or a short (see principle on Faults) and

(4) then use the "divide and eliminate" method to locate the problem (see principle on Divide & Eliminate)

<u>Divide & eliminate = 3</u>

(1) When there are no symptoms to provide any hints about the location of the fault, a systematic approach to locating the problem is helpful. One approach is to use a "divide and eliminate" method, where you section the circuit into areas and test at the point that divides one area from another (dividing points) to eliminate parts of the circuit. This allows you to systematically shrink the problem area step by step and quickly figure out where the fault most probably is.

(2) The circuit is divided into five main areas:

1. Input (supply) area (TB1 to the fuse)
2. ON-buttons (PB1, PB2, PB3) area
3. OFF-buttons (PB4, PB5, PB6) area
4. Relay (R1) area
5. Lights (LT1, LT2) area

(3) If the fault can be in any of the four main areas below the fuse, the first logical division of the circuit to divide it in two: (1) Button area and (2) Relay and lights area. In addition there is the return (neutral) path (from LT2-N to TB1-N). There is never any need to test the return path in this simulation.

<u>Voltmeter = 5</u>

(1) The Voltmeter is used to measure the voltage in the circuit and should be used on a live (energized) circuit (the current is switched on).

(2) Voltmeter is the best tool to detect an open in the circuit because there should be voltage present before the open and not after (note however, that the whole circuit is only energized when the ON-buttons are pressed)

(3) To get valid readings when using the Voltmeter, the black lead should always be on the reference point or ground (TB1-G) while the red lead should be put on the testing point.

(4) Interpreting Voltmeter readings: The circuit receives 115.0 VAC from an external circuit breaker. The only readings you will see on the Voltmeter are either 115.0 V (the part being tested is energized/receives voltage) or 0.0 V (the part being tested is not energized/doesn't receive voltage).

(5) When the ON-buttons are not pressed you should expect voltmeter readings of 0.0 V below the ON-buttons. The only exception is the relay (R1) – see information on relay for more details. When the ON-buttons are pressed all the circuit is energized and you should expect voltmeter readings of 115.0 V in the circuit (with an exception of the return path, from LT2-N to TB1-N, and the relay).

<u>Ohmmeter = 8</u>

(1) The Ohmmeter is used to measure the resistance of components in the circuit and should never be used on an energized (live) circuit because of shock hazard. You should always make sure that the current has been turned off and locked out before using the Ohmmeter (if you don't you will receive a safety error and have to start the task again).

(2) Ohmmeter is the best tool to find a short in the circuit

(3) When using the Ohmmeter, the black lead does not have to be on the reference point or ground (TB1-G) – but usually that is the most useful place for it. The red lead should always be put on an open point in the circuit which can be created, for example, by removing the fuse or disconnecting wires.

(4) Locating the probable cause: This is best done by creating an open by removing the fuse and then keeping the leads stationary (red on FU-2 and black on ground) while you press the buttons and use the screwdriver to locate the problem. You can see how the resistance changes (and whether or not the short is visible) depending on which part of the circuit is disconnected.

(5) If the short is immediately visible (indicated by a reading of low resistance before you press the ON-buttons or unscrew anything) the short has to be in the ON-buttons. If the short is only visible when the ON-buttons are pressed the short has to be in the circuit areas below the ON-buttons. This is because the ON-buttons are always open and the circuit is not completed unless one of them is pressed.

(6) Use the screwdriver to open test points to further locate the short (remember to only have one test point open at a time). If the short is visible then it must be above the test point, but if it is not it must be below the test point. (Note: remember to press the ON-buttons when testing if the short is below the ON-buttons).

(7) Testing the probable cause: Once you think you have located the faulty component you can put the red lead on the component you want to test (terminal or wire). A low resistance reading would indicate that this component is shorted to ground, but to be sure it is good to disconnect the component from other components in the circuit (for example the other end of a wire). If the reading does not change you know this is the shorted component.

(8) Resistance readings: Low resistance reading indicates a short. For example, 0.0 or 1.2 Ohms would indicate a shorted component in the circuit. Normal resistance reading is OL in all components but the relay (R1) which has a normal resistance of

1691 Ohms, and the lights which have a resistance of 26 Ohms (a reading of 13 and 52 is also normal around the lights depending on where you are testing and whether the current runs in parallel paths).

<u>Lockout = 2</u>

(1) One work practice that protects you from shock hazard when troubleshooting electrical circuits is called lockout. This is used to ensure that the circuit is isolated and locked out from all potentially hazardous energy sources.

(2) You want the circuit to be live (current on) when using the voltmeter, but when testing with the ohmmeter, unscrewing wires, or changing components you want the current to be off and locked out (the switch is sealed). You can turn off and lock out the current on the breaker panel in the simulation. If you don't remember to turn off and lock out the current before using the ohmmeter, the wrench or the screwdriver you will get a safety error (see principle on Errors), and will need to start the task again. Once you have fixed the fault and are ready to test whether it is indeed fixed, you must remember to turn the current on again.

<u>Errors = 2</u>

There are two levels of errors monitored by the system: Good practice errors and safety errors.

(1) A good practice error is less serious, and if you make one it will result in a tip being displayed urging you to use the good practice method.

- You should only have one wire unfastened at a time.
- You should open a point in the circuit before taking ohmmeter readings and take ohmmeter readings from the open point – either the wire ends or the terminal.

- You should completely remove your leads before changing meter modes.
- You should remove all leads from the circuit when using the ohmmeter before unlocking the current.

(2) A safety error is more serious, and if you make on it will result in the task shutting down and you will need to start it again.

- Never attempt to disconnect or reconnect a wire while the circuit is live
- Never attempt to replace a component while the circuit is live
- Never attempt to take ohmmeter readings on a live circuit
- Never attempt to turn on the current while the ohmmeter leads are attached
- Never attempt to remove or replace the fuse while the circuit is live
- Never attempt to replace a component while the circuit is live

<u>Replace components = 3</u>

(1) When you have determined what component is at fault you can fix the circuit by replacing the component with the wrench.

(2) Every component in the circuit can be replaced.

(3) Before you replace a component you always have to make sure that the current is turned off and locked out (see principle on Lockout), otherwise you will receive a safety error (see principle on Errors) and have to start the task again.

Open Ended Questions on the Knowledge Tests

On the two knowledge tests, half the questions (18 of the 36 questions) were open ended and needed to be coded for accuracy. Each question was worth one point, but half a point could be given for partial credit. The coding rubric was created from the principles and then three raters independently used the rubric to grade the same set of tests and, through a process of checking consistency and discussing sources of discrepancies, the key was refined. Once issues had been resolved the key was used by the three raters to grade six tests of each version of the test. An interrater reliability criterion of .8 was used (Cohen, 1988; Landis & Koch, 1977).

The absolute agreement ICC for the three raters was .874 on the X version of the test (95% CI: .573, .980; $p < .001$) and .948 for the Y version of the test (95% CI: .716, .992; $p < .001$). This indicates high interrater reliability and that the rating scale can be expected to be used effectively by different raters (McGraw & Wong, 1996; Shrout & Fleiss, 1979). The majority of the knowledge tests were graded by two of the three original raters, but the last third of the knowledge tests in the second experiment were coded by a fourth rater (recruited to replace a coder at the later stages of the experiment). After training, this rater and one of the original raters graded six of the same knowledge tests. The ICC was .813 (95% CI: .419, .939; $p < .001$) and above the adopted criterion for interrater reliability.

Rubric for Coding Open Ended Questions

Each question is worth 1 point:

- Multiple-choice questions – 1 point if correct answer selected
- Open questions – 0.5 points if partly correct, 1 if correct

<u>Version X</u>

1. Describe the divide and eliminate method. (Please be as specific as you can)

   [Divide&Eliminate 2a]

- *You section the circuit into predefined areas/halves and test at the point that divides one area from another to eliminate areas of the circuit that are working normally (not necessary to list the areas)*
- *Get 0.5 for saying sectioning the circuit; Get 0.5 for saying eliminating areas*
- *Don't get anything for saying that the method is used for locating the problem*

2. Where would you measure to check if the relay is receiving voltage? Please circle a single testing point on the picture (omitted here). [Relay 1a]
    - *R1-2*
    - *Get 0.5 if say you can see if on whether the point in the middle fills up*
3. You are troubleshooting a fault and you have figured out that the fuse is blown and when you test with the ohmmeter (black on ground and red on FU-2 with the fuse removed) you get a reading of OL by default but it changes to 0.0 when you press any of the ON-buttons. What can you conclude from this? (Please be as specific as you can). [Ohmmeter 5a]
    - *There is a short below the ON-buttons*
    - *It is not correct to say that the short might in some component or not – you don't know that – the only known thing is that it is NOT in the ON-buttons*
    - *Also not correct to say that the short is below the testing site – it usually always is when testing from this point – they need to be more specific here to get it right*
    - *Gets 0.5 for saying there is a short*

4. When lights are **on**, what voltmeter reading should you expect at the terminal before the first ON-button (PB1-1) when the circuit is operating normally? [Voltmeter 1a]
    - *115 V*
5. You have figured out that the wire between PB2 and PB3 is responsible for the circuit not working. What do you do next? [Replace components 1a]
    - *Replace the wire*
6. What is the best tool for finding an open in the circuit? [Faults 2a]
    - *Voltmeter*
7. What is the best approach for locating the problem when you don't have any symptoms to help you? (Please be as specific as you can). [Symptoms 3a]
    - *Divide & Eliminate: Divide the circuit into areas/two halves and eliminate one by one systematically by testing at the dividing point (not necessary to list areas and would give full for just saying "divide and eliminate" without further explanation)*
8. When using the ohmmeter what do you have to do before placing the red lead (you have already locked out the current and placed the black lead)? [Ohmmeter 4a]
    - *Open a point in the circuit*
    - *Give 0.5 for saying remove the fuse (because that is one way of opening the circuit)*
    - *Before changed question: Get 0.5 for saying they need to turn off and lock out current*

9. Why is it important to turn off and lockout the current? [Lockout 2a]
    - *It protects you from shock hazard and ensures that the circuit is isolated from all potentially hazardous energy sources*
    - *Get 0.5 for saying it is a safety issue*
10. When lights are **on**, what reading should you expect at the terminal before the first light (LT1-1) when the circuit is operating normally? [Voltmeter 2a]
    - *115 V*
11. What does the ohmmeter measure and what kind of fault is it best for detecting? [Ohmmeter 3a]
    - *Resistance, Short*
12. You press the buttons on the circuit and notice that the relay gets energized when you press the ON-buttons and de-energized when you press the OFF-buttons, but the lights are dead. What can you conclude about the location of the problem? [Symptoms 2a]
    - *You know that the relay is receiving current and the problem must be in the relay or the lights*
    - *Get 0.5 if only say that the problem is in the lights*
13. What is a short? [Faults 1a]
    - *When two isolated components come into contact and the current travels on an alternate path (for example wires get frayed and come into contact with ground)*

14. If there is a short in the wire from PB2 to PB3: (a) how would the circuit behave? (b) What meter readings would you expect to see when measuring at FU-2 and pressing the buttons? Please be as specific as you can. [Tasks 2a]
    - *a: The fuse would be blown, the relay wouldn't get energized and the lights would not turn on (get right for saying the circuit wouldn't work*
    - *b: using the ohmmeter I would expect 0.0 at FU-2 (whether or not the buttons are pressed*
15. What does the state of the fuse tell you about the problem in the circuit? [Fuse 2a]
    - *If the fuse is blown then it's a short (0.5), if it's not blown then it's an open (0.5)*
    - *Note: In either case the problem is not just the fuse, but the state of the fuse will indicate whether to use the voltmeter to look for an open or the ohmmeter to look for a short (a blown fuse is a sign that there is a short in the circuit somewhere)*
16. You know you are looking for an open in the circuit and you have narrowed the possible location down to the button area. What do you do next to find the open? Please be as specific as you can. [Tasks 3a]
    - *Divide the button in two areas (ON and OFF) and use the voltmeter to eliminate either area by testing at the dividing point*
17. What happens if you press one of the OFF-buttons (when the lights are **on** and the circuit is behaving normally)? Please be as specific as you can. [Circuit 2a]

- *The normally closed OFF button becomes open temporarily, relay disengages, and lights turn off*
- *0.5 if only says the lights turn off + 0.5 is say either of the other*

18. How do you measure voltage and when is it important to do so? [Voltmeter 4a]
    - *You measure voltage with a voltmeter (0.5), and it's important to do so when you have an open in the circuit (0.5)*

Version Y

1. You know you are looking for a short in the circuit and you have narrowed it down to either the OFF-buttons or the relay. What do you do next to find the short? Please be as specific as you can. [Tasks 3b]
    - *Divide the area in half – OFF buttons and relay – and test with the Ohmmeter at the dividing point to eliminate either area.*
    - *Saying that you should check everything within those areas or using an ohmmeter to find the short does not get you any points*
2. What is an open? [Faults 1b]
    - *A break in the circuit that prevents current from flowing*
    - *The circuit cannot be completed is also correct*
3. What does the voltmeter measure and what kind of fault is it best for detecting? [Voltmeter 4b]
    - *Voltmeter measures voltage (0.5), and it's best for detecting an open (0.5)*
    - *Correct to say that the voltmeter measures charge*
    - *Incorrect to say the voltmeter measures watts or level of energy or electricity*

4. You have figured out that the relay is responsible for the circuit not working. What do you do next?   [Replace components 1b]
    - *Replace the relay]*
5. If there is an open in the wire between the two lights, how would the circuit behave? [Symptoms 2b]
    - *The first light would turn on, but the second wouldn't – get full points for mentioning only the 2nd light*
    - *Saying the relay dot fills up is irrelevant*
6. What is the best tool for finding a short in the circuit? [Faults 2b]
    - *Ohmmeter*
7. You are troubleshooting a fault and you have figured out that the fuse is blown and when you test with the ohmmeter (black on ground and red on FU-2 with the fuse removed) you get a reading of 0.0 by default (without pressing any of the buttons). What can you conclude from this? (Please be as specific as you can). [Ohmmeter 5b]
    - *There is a short in the ON-buttons*
    - *Gets 0.5 for saying there is a short*
8. When lights are **off**, what reading should you expect at the terminal before the first ON-button (PB1-1) when the circuit is operating normally? [Voltmeter 1b]
    - *115 V*
9. You start troubleshooting a new fault and nothing happens when you press the ON-buttons. What would you do next? (Please be as specific as you can). [Symptoms 3b]

- *Check the fuse to see if the problem is an open or short (0.5); Divide and eliminate (0.5)*

10. What safety measure protects you from shock hazard? [Lockout 2b]
    - *Turning off (0.5) and locking out the current (0.5)*
11. How do you measure resistance and when is it important to do so? [Ohmmeter 3b]
    - *With an ohmmeter, when you have a short in the circuit*
12. When lights are **off**, what voltmeter reading should you expect at the terminal before the first light (LT1-1) when the circuit is operating normally? [Voltmeter 2b]
    - *0 V*
13. What is the purpose of removing the fuse when testing with the ohmmeter? [Ohmmeter 4b]
    - *To create an open in the circuit to test from*
14. What happens if you press one of the ON-buttons (when the lights are **off** and the circuit is behaving normally)? Please be as specific as you can. [Circuit 2b]
    - *Relay gets energized, seal-in contacts close, and lights turn on*
    - *Get 0.5 for saying lights turn on and +0.5 for saying either of the other two*
15. Why is it important to investigate further when you find the fuse blown? [Fuse 2b]
    - *Because a blown fuse is a symptom of a short somewhere in the circuit*
    - *For saying a problem in the circuit is a reason why the fuse blew (0.5)*

16. If there is an open in either relay contact: (a) would the relay dot fill up (the relay get energized)? (b) What meter readings would you expect to see when measuring at R1-6? [Tasks 2b]

    - *a: Yes, the relay would get energized – the dot would fill up (0.5)*
    - *b: 0 V (0.5)*

17. Where would you measure to check if the relay is providing the lights with voltage? Please circle a single testing point on the picture (omitted here). [Relay 1b]

    - *R1-6*

18. How does the divide and eliminate method work? (Please be as specific as you can). [Divide&Eliminate 2b]

    - *You section the circuit into predefined areas/halves and test at the point that divides one area from another to eliminate areas of the circuit that are working normally (not necessary to list the areas)*
    - *Get 0.5 for saying sectioning the circuit; Get 0.5 for saying eliminating areas*
    - *Don't get anything for saying that the method is used for locating the problem*

Drawings of the Circuit

After completing the testing tasks in the second session the participants were asked to draw the circuit from memory in as much detail as they could.

The main focus of the coding rubric was whether the participant had depicted all the elements (components and wires) correctly, then on details important for function (in terms of locating faults and finishing tasks), and lastly on details not relevant for functionality. Points were given for showing the structure of the circuit (structure points), depicting functional details (functional detail points), and including general details in the drawing (drawing detail points).

Structure points were given if a component or a wire was depicted (i.e., 2 points were given for a component such as buttons, relay, and fuse, but 1 point was given for a single wire). Functional detail points were given if certain details important for indicating the function of the circuit were depicted (e.g., by drawing the wires between the ON-buttons accurately). These functional details are important for diagnosing faults and completing tasks. Drawing detail points were given for providing details of the circuit, such as specifying names of components and drawing wires not crucial to the diagnosis of the faults (e.g., neutral wires). These details show good memory for how the circuit looks but are not necessary to indicate functionality. Separating the points for structure, function, and drawing details was done because even though participants were instructed to include as much details as possible, it is likely that participants would differ in terms of how much time they were willing to devote to drawing the circuit and what they consider adequate details. That is, one cannot assume that fewer details in the drawing indicate poorer recall or mental model of the circuit.

Participant could at most get 39 points (26 for structure, 7 for functional details, and 6 for drawing details). Two raters coded 10 of the same drawings and interrater reliability was calculated with ICC and .8 criterion was adopted (Cohen, 1988; Landis &

Koch, 1977). The absolute agreement ICC between the two raters on the total scores was .97 (95% CI: .89, .99; $p < .001$), indicating a high interrater reliability.

Rubric for Coding Drawings

Total of 39 points

<u>Components = 22 points</u>

Structure (is the component there?) = 14 points

- TB1 = 2 p
- Fuse = 2p
- PB1 / PB4 = 2p
- PB2 / PB5 = 2p
- PB3 / PB6 = 2p
- R1 = 2p
- Lights (L1/L2) = 2p

Functional details (what does it do?) = 4 points

- TB1:
    - Has three parts = 1p
    - Parts named (at least G) = 1p
- R1:
    - Draws all eight terminal points = 2p

Drawing details = 4 points

- Fuse:
    - Names the component (either fuse or FU) = 0.5p

- Buttons:
    - Which is on and which is off (needs to specify for all) = 0.5p
    - Name of button (e.g. PB1, PB4; needs all) = 1p
- R1:
    - Names the component (either relay or R1) = 0.5p
    - Names terminal points – at least four (e.g., R1-1) = 1p
- Lights (L1/L2):
    - Names the components = 0.5p

<u>Wires = 17 points</u>

Structure (is the connection there?) = 1 point per connection for a total of 12 points

- TB1 – FU
- FU – PB1
- PB1 – PB2 (only need to show the connection, even if there are two wires)
- PB2 – PB3 (only need to show the connection, even if there are two wires)
- PB3 – PB6
- PB6 – PB5
- PB5 – PB4
- PB4 – R1
- PB6 – R1
- R1 – LT1
- LT1 – LT2
- R1 – FU

[Note: The main point is to show the connection between components, so both end of wire have to be connected to the right components.]

Functional details = 3 points

- ON-Buttons:
    - Wires between terminals 1-1 (two wires: PB1-PB2, PB2-PB3) = 1p
    - Wires between terminals 2-2 (two wires: PB1-PB2, PB2-PB3) = 1p
    - [Note: give 0.5 point for each correct]
- LT1 – LT2:
    - Shows both wires accurately (terminals L-L and N-N) = 1p

Drawing details = 2 points

- TB1: Shows connection out of circuit = 0.5p
- Wire R1-1 – R1-8 = 0.5p
- Wire LT2-N – R1 (neutral return wire) = 0.5p
- Wire R1 – TB1 (neutral return wire) = 0.5p

Explanations of the Circuit

At the end of the delayed testing session the participants asked to explain the workings of the circuit with three questions. The first question showed a picture of the circuit and participants were asked to provide a short explanation of the function of each component. The coding rubric for this question was based on the principles and each component (TB1, fuse, relay) or type of component (ON-buttons, OFF-buttons) could give a certain number of points depending on the informational content of the principle in question. In addition, more points were given for advanced knowledge than basic knowledge. For example, the information that the ON-buttons energize the relay and turn the lights on gives one point (basic knowledge), but the information that the ON-buttons are by default open gives two points (advanced knowledge). A participant could get a maximum of 27 points for this question.

The second question showed a close up of the relay and the participants were asked to use the picture as reference and provide as much detail about how the relay works as possible (specifically the relay coil and contacts). This question was added because pilot testing had indicated that the relay was the component participants found the most complex and difficult to understand. Therefore, the understanding of this component could separate those with good knowledge of how the circuit works from those with less knowledge. This question was coded using the principle about the relay and how the circuit works as guidelines. Participants could get four points for describing how the relay worked in general (e.g., relay keeps the lights energized or de-energized after the buttons are released), three points for describing the coil (e.g., relay coil receives voltage through R1-2 and a reading at that point will show whether or not the relay coil is receiving voltage), and 7 points for describing the seal-in contacts (e.g., the two points of the seal-in contacts are never energized at the same time). There was a maximum of 22 points given for this question.

The third question showed a picture of the circuit and the participants were asked to use a red pen to mark where the circuit is energized when the lights are off (assuming the circuit is functioning properly). They were given one point for each correct part highlighted (e.g., wire from TB1 to fuse) for a maximum of seven points, but were deducted one point for incorrectly marked parts (zero was the lowest possible score as negative points were not given). The deduction was used so that someone coloring every single wire and component in the circuit would not get full points; instead the full points were given for marking only the correct parts, nothing more and nothing less.

The same rater coded all the explaining questions assuring some level of reliability. However, to make sure the coding rubric was objective and could be reliably used by other coders if necessary, a second rater also coded a random sample of 12 participants. An interrater reliability criterion of .8 was adopted (Cohen, 1988; Landis & Koch, 1977). The absolute agreement *ICC* for the two raters was .84 for the first question (95% *CI*: .55, .95; $p < .001$), showing good interrater reliability and that the rubric could reliably be used by a different rater with similar results.

For the second question the ICC was .70 (95% CI: -.16, .92; $p < .05$) and below criterion. A comparison of the two raters showed that one was consistently stricter than the other, giving lower scores across the board. But, because the raters had a good agreement in rank ordering of the participants ($r = .83$) and ultimately only one of the raters rated all the tests a reliability of .7 was considered good enough in this case.

Coding for the third question involved counting components and wires and therefore contained less subjective interpretations than the first two questions. But, because sometimes participants markings were unclear (e.g., one end of wire was highlighted, or a whole component circled) ICC was calculated. It was found to be acceptable (.91, 95% CI: .50, .95; $p < .001$).

Rubric for Coding the Explanation Questions

<u>Question #1 = 27 points</u>

TB1 = 2 points

a) Supplies power = 1 p

b) Ground (when using multi-meter) = 1p

Fuse = 6 points

Basic

a) The fuse is designed to protect the circuit (when too much current flows through it) = 1 p

b) An excessive current causes the fuse to open (blows) = 1 p

Advanced

c) An open interrupts the flow of current and isolates the circuit (protecting components that could otherwise be damaged) = 2 p

d) Typically a short to ground causes excessive current / If you find a blown fuse chances are there is a short to ground in the circuit = 2 p

ON-Buttons (PB1-3) = 6 points

Basic

a) Pressing any of the ON-buttons (PB1, PB2, or PB3) energizes the lights – turns the lights on = 1 p

b) Pressing any of the ON-buttons (PB1, PB2, or PB3) energizes the relay (R1) = 1 p

Advanced

c) The ON-buttons are by default open (create a break in the path of the current) = 2 p

d) When ON-buttons are pressed they close and complete the circuit (energizing the relay and the lights) = 2 p

OFF-Buttons (PB4-6) = 6 points

Basic

a) Pressing any of the OFF-buttons (PB4, PB5, or PB6) de-energizes the lights and turns the lights off = 1 p

b) Pressing any of the OFF-buttons (PB4, PB5, or PB6) de-energizes the relay = 1 p

Advanced

c) The OFF-buttons are by default closed (meaning that they complete the circuit unless they are pressed) = 2 p

d) When an OFF-button is pressed it creates an open in the circuit (de-energizing the relay and the lights) = 2 p

Relay (R1) = 7 points

Basic

a) The dot in the center of the relay lights up when the relay is working = 1 p

b) Energizing the relay provides current to the lights, de-energizing the relay prevents current from flowing to the lights = 1 p

c) The relay keeps the lights either energized or de-energized after the buttons are released = 1 p

Advanced

d) When the relay gets energized the two seal-in contacts in the relay close (energizing the lights) and stay close even when the ON-button is released (becomes open again) = 2 p

e) When the relay is de-energized the two seal-in contacts will open (de-energizing the lights) and stay open even when the OFF-button is released (becomes closed again) = 2 p

NOTES

- Not give any points for describing what the lights do (turn on and off) – that should be self-evident
- Not necessary to say what the wires do
- When testing the rubric take note of whether someone you think is good and thorough gets a high score or not and someone you think does a lousy job gets a low score or not. If the score do not make this distinction we need to change them somehow.
- Take notes when starting to grade – we'll use that when reevaluating the scoring scheme.

<u>Question #2 = 22 points</u>

General = 6 points

Basic

a) Energizing the relay provides current to the lights, de-energizing the relay prevents current from flowing to the lights = 1 p

b) The relay keeps the lights either energized or de-energized after the buttons are released = 1 p

Advanced

c) When the relay gets energized the two seal-in contacts in the relay close (energizing the lights) and stay close even when the ON-button is released (becomes open again) = 2 p

d) When the relay is de-energized the two seal-in contacts will open (de-energizing the lights) and stay open even when the OFF-button is released (becomes closed again) = 2 p

Coil = 6 points

a) The relay coil receives voltage through R1-2 and this means that a reading at R1-2 will show whether the relay coil is receiving voltage (can also be seen by whether the dot in the middle of the relay fills up or not) = 2 p

b) The return path from the relay coil is R1-7. In the tasks you will do, it is never going to be informative to test at R1-7 because the voltmeter reading at R1-7 is always 0 V = 2 p

c) Even if the relay coil is receiving current (and gets energized) the problem can still be in the relay's seal-in contacts = 2 p

Seal-in contacts = 10 points

Basic

d) The two seal-in contacts receive voltage at R1-1 and R1-8 = 1 p

e) These points (R1-1 and R1-8) are always energized if the fuse is working = 1 p

f) The first seal-in contact is R1-3 and R1-4 = 1 p

g) The second seal-in contact is R1-5 and R1-6 = 1 p

Advanced

h) If a seal-in contact is open (not letting current through the relay) R1-4/R1-5 will be energized (but not R1-3/R1-6) = 2 p

i) If the contact is closed (letting current through the relay) R1-3/R1-6 will be energized (and not R1-4/R1-5) = 2 p

j) The two points of the seal-in contact are therefore never energized at the same time = 2 p

<u>Question #3 = 7 points</u>

When the lights are off (under normal operating conditions) the following parts are energized:

- Wire from TB1-L to FU-1 (supply to fuse)
- Wire from FU-2 to PB1-1 (fuse to on-buttons)
- Wire from PB1-1 to PB2-1 (on-button 1 to on-button 2)
- Wire from PB2-1 to PB3-1 (on-button 2 to on-button 3)
- Wire from FU-2 to R1-1 (fuse to relay)
- Wire from R1-1 to R1-8 (within relay)
- Open seal-in contacts R1-4 and R1-5 (within relay)

= Get 1 point for each correct

= Get 1 point deducted for incorrect

Note: Cannot get less than zero points!

- Can either highlight the wires or the terminals (screws)
- Deduct 1 point for incorrect wire
- Deduct 0.5 point for incorrect terminal

APPENDIX F

Knowledge of Fault Results

Experiment 1

Knowledge of Fault in Training

After completing each task, participants were asked to describe what was wrong with the circuit and what they did to fix it. The answers were coded as either correct or incorrect and those participants who used the third hint to complete the task were excluded from the analysis because the third hint states explicitly what is wrong with the circuit (on average 13% of the training tasks were completed with the help of the third hint and there was no difference among the groups, $p > .05$).

Figure 22 shows the percentage of correctness for the participants in each condition. The Read-Before generally had the most correct answers and the Summarize-After the fewest. The difference among the groups was not significant ($p > .05$).

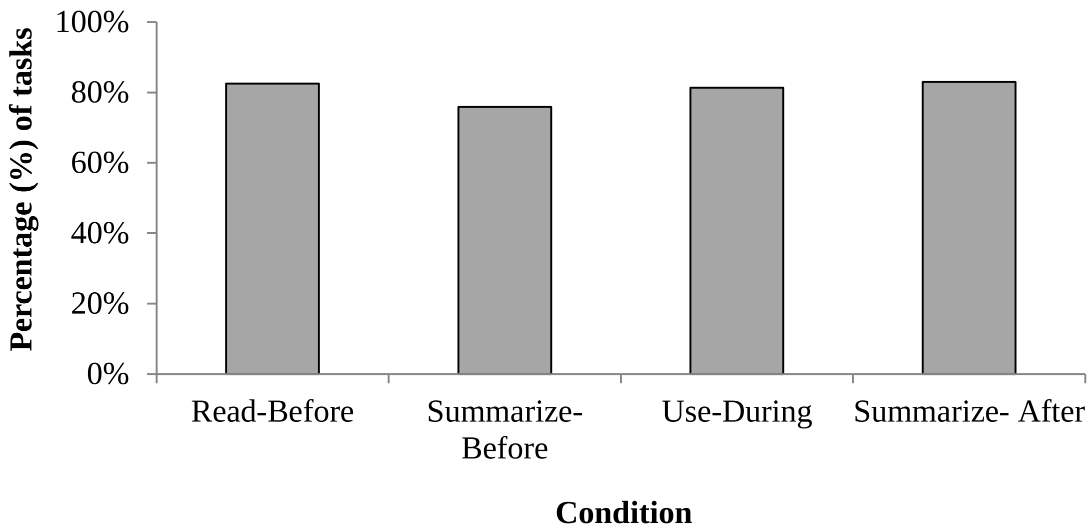


Figure 22. Percentage of training tasks in each condition where participants correctly described what was wrong with the circuit after fixing it. The cases where participants had used the last hint to complete tasks were excluded because the third hint explicitly stated what was wrong with the circuit.

<u>Knowledge of Fault in Testing</u>

After each testing task participants were asked to describe what was wrong with the circuit and what they did to fix it. These responses were rated for correctness, and I proposed that the participants in the Use-During group would provide more correct descriptions than the participants in the summarizing groups. The descriptions of the fault were coded either as incorrect (did not identify the component that had to be replaced) or correct (identified the component that had to be replaced).

All the occasions where participants had used the third hint to complete the task were excluded because the third hint spelled out exactly what component needed to be replaced. The participants could therefore repeat this information when asked what the fault was even if they had not successfully completed the task.

In about 80% of tasks, the participants in each condition correctly identified the faulty component after completing the testing tasks (see Figure 23), and there was no difference among the conditions ($p > .05$).

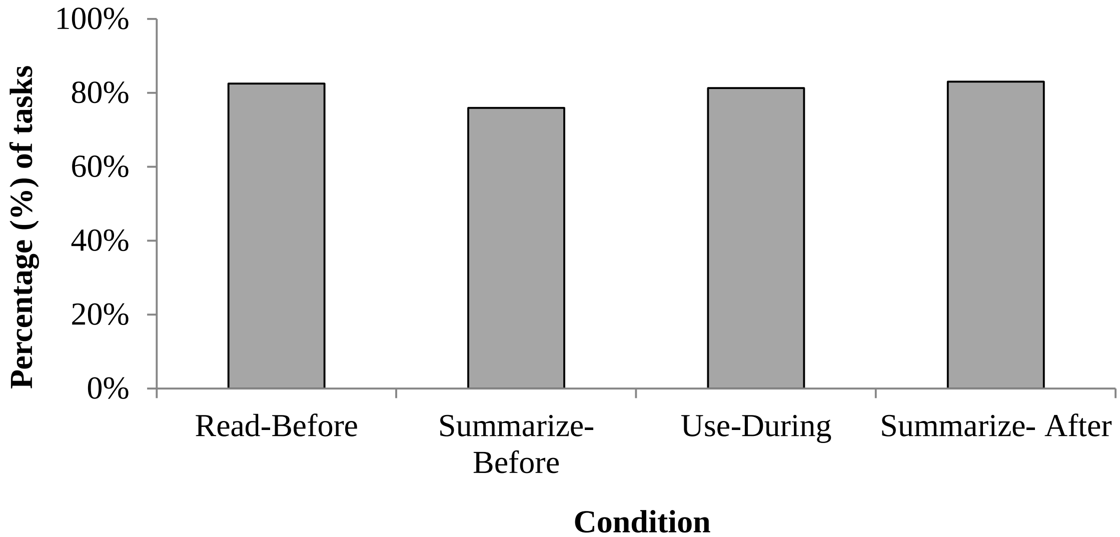


Figure 23. Percentage of testing tasks in each condition where participants correctly identified the faulty component.

Interestingly, many participants who were not able to identify the faulty component stated that they had changed many if not all the components in the circuit and therefore had no idea which one was at fault. This indicates that the number of unnecessary components replaced might be a good indicator of problem solving proficiency, at least separating those adopting the strategy of replacing all the components from all the others.

Experiment 2

Knowledge of Fault in Training

Once participants had completed each training task, they were asked to describe what had been wrong with the circuit and what they did to fix it. These descriptions were coded as either correct or incorrect based on whether they had correctly identified the faulty component. Tasks where participants had used the third hint to help them were excluded from the analysis because they were told exactly what component was faulty. On average 9% of the training tasks were completed with the help of the third hint, and the participants using the general procedural instructions were more likely to use the third hint (15%) than participants using the detailed procedural instructions (3%; $p < .05$).

Figure 24 shows the percentage of tasks where participants correctly identified the faulty component. For the majority training tasks in all four groups the participants knew what had been wrong with the circuit (87-99%). The Use-During-Detailed group had the highest percentage of correctly identified faults, with 99% describing the fault correctly. However, neither timing of principle use nor procedural instruction specificity, influenced knowledge of fault ($p > .05$).

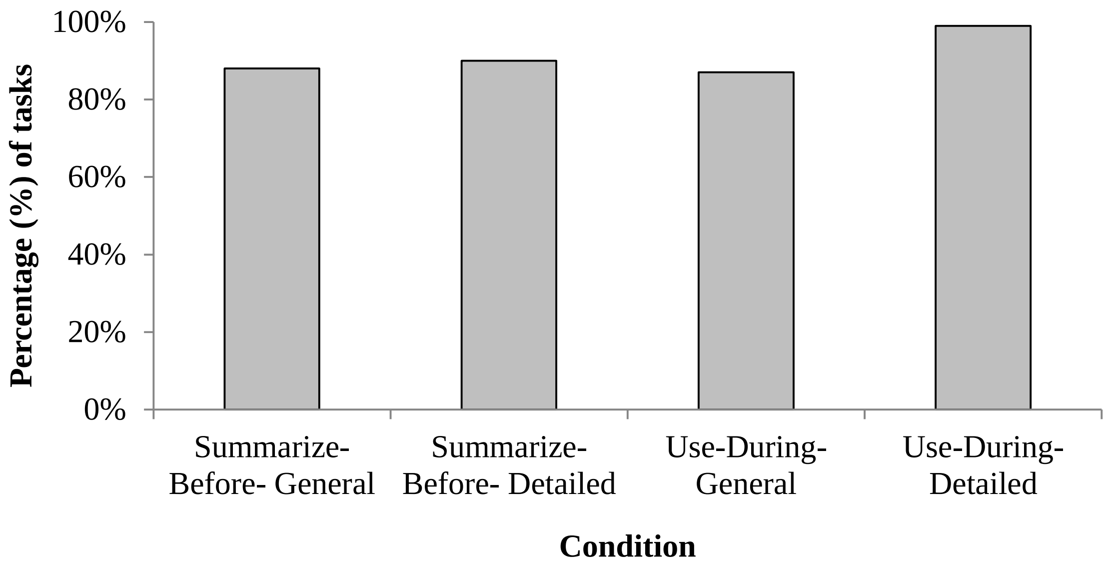


Figure 24. Percentage of training tasks in each condition where participants correctly described what was wrong with the circuit after fixing it. The cases where participants had used the last hint to complete tasks were excluded because the third hint explicitly stated what was wrong with the circuit.

<u>Knowledge of Fault in Testing</u>

On task completion, participants were asked to describe what the fault had been and what they had done to fix it. The descriptions were rated for correctness depending on whether the participants correctly identified the component that had to be replaced to fix the circuit.

Participants who had needed the third hint to complete a task were excluded from the analysis because the third hint specified what component needed to be replaced, and participants could therefore show good knowledge of fault after completing the task, even if they had not been successful at doing so.

For the majority of the testing tasks the participants correctly identified what had been wrong with the circuit (see Figure 25). The percentage was a bit higher for the Summarize-Before-General group (91%) than the other three groups (83-85%), but there was no difference among the conditions based on the procedural instructions specificity or timing of instruction use ($p > .05$ in both cases). The hypothesis was therefore not supported in this case.

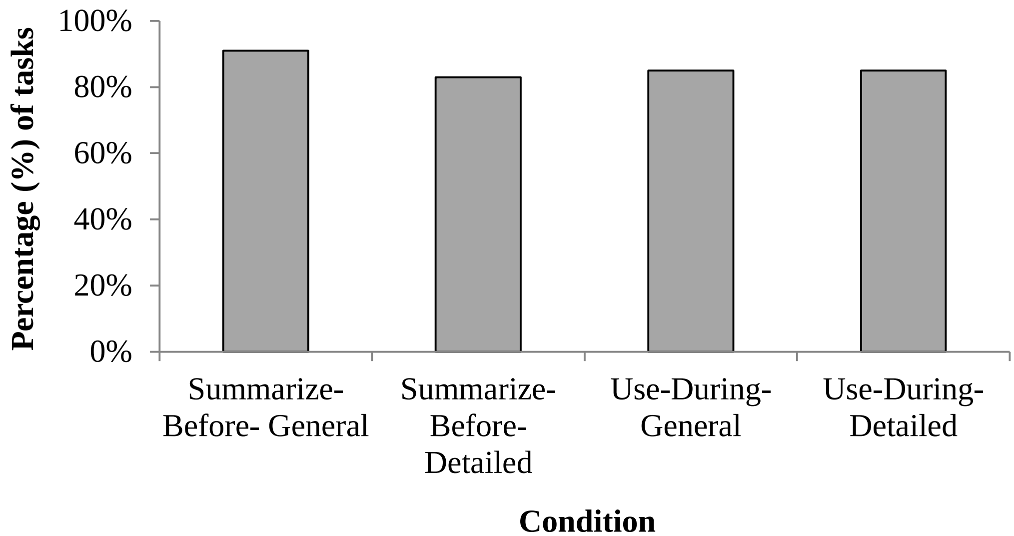


Figure 25. Percentage of testing tasks in each condition where participants correctly identified the faulty component.

APPENDIX G

Task Difficulty Results

Experiment 1

After completing each task the participants rated how difficult they found the task on a rating scale from 0 to 100, with a higher number representing more difficulty. Figure 26 shows the average difficulty ratings of the training and testing tasks for the four conditions. There averages are very similar for the groups, with the Use-During groups showing a slightly higher difficulty rating for both training and testing tasks, but there was no significant differences among the groups (p > .05 in both cases).

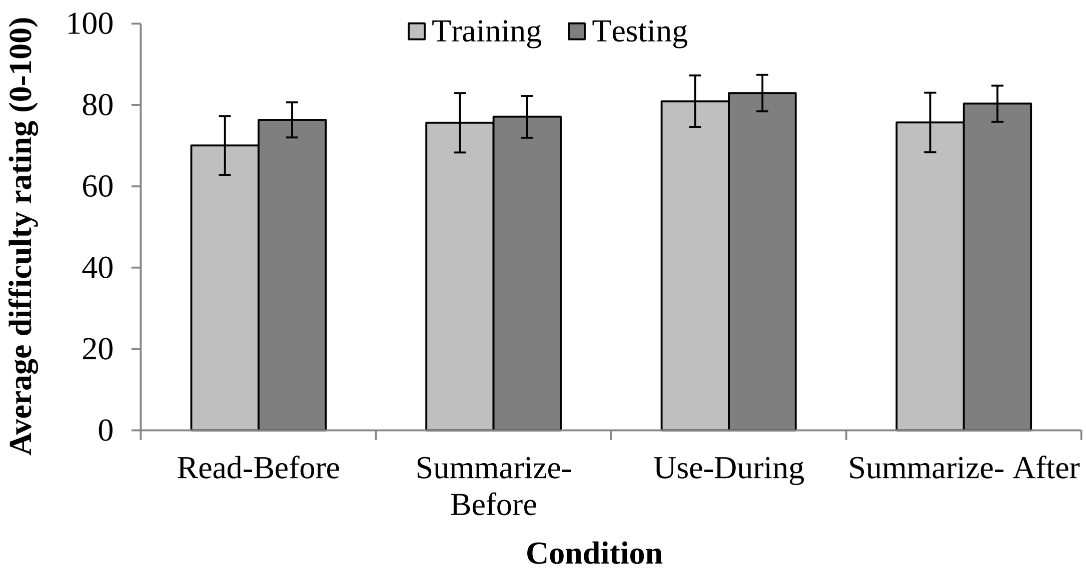


Figure 26. Average difficulty rating for training and testing tasks in each condition. The error bars show twice the standard error.

There was not a significant difference between difficulty ratings during training and testing ($p > .05$), but there was a difference for ratings of retention and transfer tasks ($t(629) = 2.07$, $p < .05$, $d = 0.17$, 95% CI [.25, 9.39], with retention tasks being rated as more difficult ($M = 81.84$, $SD = 27.34$) than the transfer tasks ($M = 77.02$, $SD = 30.08$). This contradicts earlier findings showing worse performance on the transfer task compared to the retention tasks. There was no interaction between condition and experiment part (training vs. testing, $p > .05$) or condition and test type (retention vs. transfer, $p > .05$).

In conclusion: The task difficulty rating did not yield any differences among the groups, for training or testing, and the hypothesis was therefore not supported in this case. Also, no differences were found between difficulty ratings of tasks completed during training and testing. There was a significant difference in difficulty ratings for retention and transfer testing tasks, but in the opposite direction of what might be expected as the retention tasks were rated more difficult. Because of these unexpected results, I looked back at the raw data to see how participants were rating the tasks and found that a number of participants (19%) rated most or all tasks at ceiling, and therefore had very little variation in difficulty ratings. The difficulty rating is therefore shows a ceiling effect and is not a particularly informative measure in this context.

Experiment 2

After completing each task the participants rated how difficult they found the task on a rating scale from 0 to 100, with a higher number representing more difficulty. Figure 27 shows the average difficulty ratings of the training and testing tasks for the four conditions. The averages were overall very high (around 80), both in training and testing.

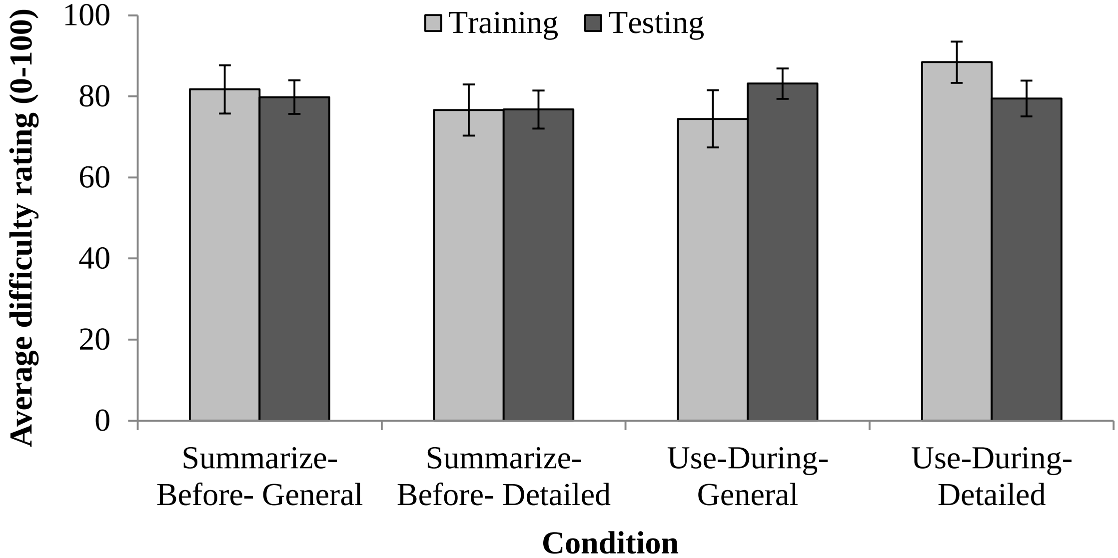


Figure 27. Average difficulty rating for training and testing tasks in each condition. The error bars show twice the standard error.

There was an interaction between timing of principle use and procedural instruction specificity for the difficulty ratings during training, $F(1,281) = 9.75$, $p < .05$, $\eta_p^2 = .03$. The interaction was followed up with simple effect analyses for each level of timing of principle use. For the participants in the Summarize-Before groups there was no difference in difficulty rating based on whether they used general or detailed procedural instructions ($p > .05$), but for the participants in the Use-During groups the ones using general procedural instructions rated the tasks less difficult than the ones using detailed ($F(1,139) = 10.49$, $p < .05$, $\eta_p^2 = .07$). This is the opposite of what had been predicted, as the difficulty rating in training was higher for participants in the Use-During groups using detailed procedural instructions, but was expected to be lower than for those using general procedural instructions.

There was no effect of either timing of principle use or procedural instruction specificity on the difficulty ratings during testing ($p > .05$).

There was not a significant difference between difficulty ratings during training and testing ($p > .05$) and there was no difference between difficulty ratings for retention and transfer testing tasks ($p > .05$).

APPENDIX H

Role of Memory in Completing Retention Tasks Results

Experiment 1

For the retention tasks (testing tasks that had already been completed during training) the participants were asked whether they had remembered how they had solved the task during training and this helped them complete it. I expected that in more cases the participants in the Use-During group would remember how to solve the retention tasks than participants in the summarizing groups. In addition, I expected that generally participants who relied on memory would have a better retention performance (faster, with less safety errors, fewer unnecessary components changed, and fewer meter readings) than those who solved the task again. Each participant completed three retention tasks, one in the immediate testing session and two in the delayed testing session.

The trend in percentages of cases where the participants reporting relying on memory fell along the lines expected: Participants in the Use-During condition relied on memory to solve the retention tasks in higher percentage of cases compared to the two summarizing groups (see Figure 28), but this difference was not significant ($p > .05$). Therefore, the part of the hypothesis stating that the Use-During group would be more likely to remember the task solution was not supported.

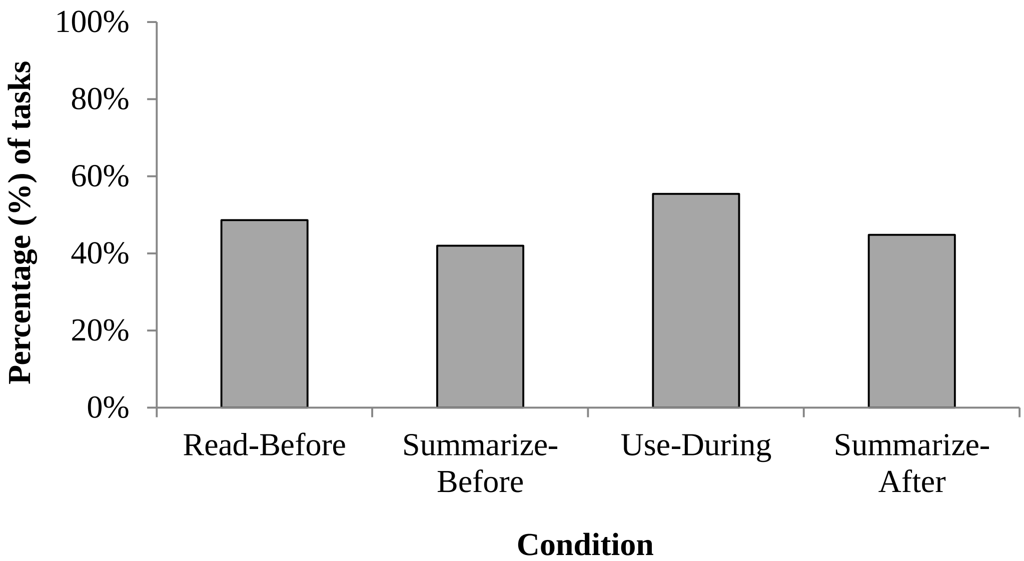


Figure 28. The percentage of retention tasks in each condition where the participants relied on memory of having completed the tasks in training.

Even if the groups did not differ in terms of whether they relied on memory when solving the retention task, the effects of relying on memory – regardless of condition – was considered for the measure of procedural learning. There was a significant effect of memory for the measure of procedural learning: Participants performed better when completing the testing tasks while relying on memory of completing the task during training (M = -0.29, SD = 0.59) than when they did not (M = 0.98, SD = 0.79), $t(269) = -4.63$, $p < .001$, $d = -0.52$, 95% CI [-0.56, -0.23].

The hypothesis that more participants in the Use-During group would remember how to solve the retention tasks as compared to the participants in the summarizing group was not supported. However, the participants who relied on memory of the retention tasks showed better troubleshooting performance than the ones who did not.

The results show that it is important to make note of whether participants rely on memory for tasks when measuring performance on task retention and making the task memorable during training should lead to better performance on that same task later.

Experiment 2

After completing the retention tasks the participants were asked whether they had relied on memory of completing the task in training when doing so. Every participant completed three retention tasks and they were explicitly told they had already completed the tasks in training.

Participants relied on memory over half the time when completing the retention tasks (see Figure 29). The participants in the Summarize-Before-Detailed group relied less on memory than the other three groups, but there was no significant effect found for either manipulation (*p* > .05 in both cases). The hypothesis was therefore not supported.

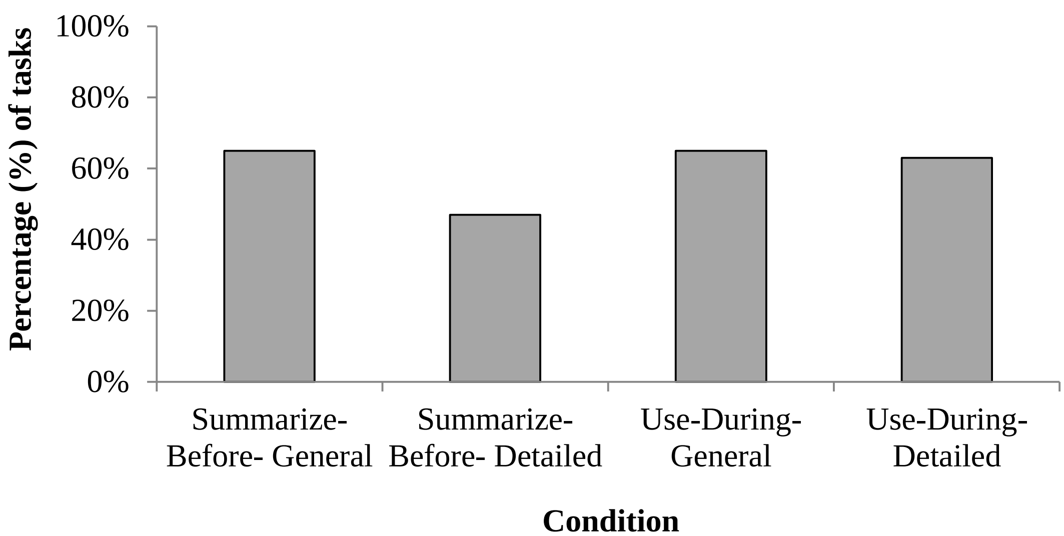


Figure 29. The percentage of retention tasks in each condition where the participants relied on memory of having completed the tasks in training.

Regardless of condition, the participants performed better on the testing tasks when they relied on memory (*M* = -0.26, *SD* = 0.66), than when they did not (*M* = 0.12, *SD* = 0.83; $F(1,189) = 11.30$, $p < .001$, $\eta_p^2 = .06$).

REFERENCES

Alterman, R., Zito-Wolf, R., & Carpenter, T. (1991). Interaction, comprehension, and instruction usage. *The Journal of the Learning Sciences, 1*(3&4), 361-398.

Andrews, D. C. (2001). *Technical Communication in the Global Community* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Ausubel, D. P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of Educational Psychology, 51*(5), 267-272.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612-637.

Berry, D., & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalized knowledge. *The Quarterly Journal of Experimental Psychology A: Human Experimental Pyschology, 36A*(2), 209-231.

Bibby, P. A., & Payne, S. J. (1993). Internalization and the use specificity of device knowledge. *Human-Computer Interaction, 8*, 25-56.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). Cambridge, MA: MIT Press.

Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Cognitive regulation of performance: Interaction of theory and application* (pp. 435-459). Cambridge, MA: MIT Press.

Borgman, C. L. (1999). The user's mental model of an information retrieval system: An experiment on a prototype online catalog. *International Journal of Human-Computer Studies, 51*(2), 435-452.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *4*(2), 717-726.

Brown, A. L., Campione, J. C., & Day, J. D. (1981). Learning to learn: On training students to learn from texts. *Educational Researcher*, *10*(2), 14-21.

Bujak, K. R. (2010). *A framework of passive-active-constructive study techniques: A divergence between assigned and reported behaviors* (Unpublished master's thesis). Georgia Institute of Technology, Atlanta, GA.

Carroll, J. M. (1987). Minimalist design for active users. In R. M. Baecker & W. A. S. Buxton (Eds.), *Readings in Human-Computer Interaction: A Multidisciplinary Approach* (pp. 621-626). Los Altos, CA: Morgan Kaufmann Publishers.

Carroll, J. M. (1990). *The Nurnberg funnel: Designing minimalist instruction for practical computer skill*. Cambridge, MA: MIT Press.

Carroll, J. M. (1998). *Minimalism beyond the Nurnberg funnel*. Cambridge, MA: MIT Press.

Carroll, J. M., & Mack, R. L. (1987). Learning to use a word processor: By doing, by thinking, and by knowing. In R. M. Baecker & W. A. S. Buxton (Eds.), *Readings in Human-Computer Interaction: A Multidisciplinary Approach* (pp. 278-297). Los Altos, CA: Morgan Kaufmann.

Carroll, J. M., Mack, R. L., Lewis, C. H., Grischkowsky, N. L., & Robertson, S. R. (1985). Exploring exploring a word processor. *Human-Computer Interaction, 1,* 283-307.

Catrambone, R. (1990). Specific versus general procedures in instructions. *Human-Computer Interaction*, *5*(1), 49-93.

Catrambone, R. (1995). Following instructions: Effects of principles and examples. *Journal of Experimental Psychology: Applied, 1*(3), 227-244.

Catrambone, R., & Yuasa, M. (2006). Acquisition of procedures: The effects of example elaborations and active learning exercises. *Learning and Instruction, 16*(2), 139-153.

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73-105.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145-182.

Chi, M. T. H., de Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439-477.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum.

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist, 60*(2), 170-180.

Davis, S., & Wiedenbeck, S. (1998). The effect of interaction style and training method on end user learning of software packages. *Interacting with Computers, 11*(2), 147-172.

Duff, S. C., & Barnard, P. J. (1990, August 1990). *Influencing behavior via device representation: Decreasing performance by increasing instruction.* Paper presented at the Human-Computer Interaction - INTERACT '90.

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363-406.

Farkas, D. K. (1999). The logical and rhetorical construction of procedural discourse. *Technical Communication, 46*(1), 42-54.

Fleming, N. (2009). VARK: A guide to learning styles. Retrieved November 20, 2009, from http://www.vark-learn.com/english/index.asp

Fleming, N., & Mills, C. (1992). Not another inventory, rather a catalyst for reflection. *To Improve the Academy, 11*, 137-155.

Fu, W., & Gray, W. D. (2006). Suboptimal tradeoffs in information seeking. *Cognitive Psychology, 52*(3), 195-242.

Ganier, F. (2004). Factors affecting the processing of procedural instructions: Implications for document design. *IEEE Transactions on Professional Communication, 47*(1), 15-26.

Gott, S. P., Lajoie, S. P., & Lesgold, A. (1991). Problem solving in technical domains: How mental models and metacognition affect performance. In R. F. Dillon & J. W. Pellegrino (Eds.), *Instruction: Theoretical and Applied Perspectives* (pp. 107-117). New York: Praeger.

Gray, W. D., Sims, C. R., Fu, W., & Schoelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review, 113*(3), 461-482.

Guthrie, J. T., Bennett, S., & Weber, S. (1991). Processing procedural documents: A cognitive model for following written directions. *Educational Psychology Review, 3*(3), 249-265.

Hicks, T. G., & Wierwille, W. W. (1979). Comparison of five mental workload assessment procedures in a moving-base driving simulator. *Human Factors, 21*(2), 129-143.

Jex, H. R. (1988). Properties of workload assessment techniques. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload*. Amsterdam, Holland: Elsevier Science Publishers.

Karreman, J. (2004). *Use and effect of declarative information in user instructions* (Vol. 18). Amsterdam: Rodopi.

Karreman, J., & Steehouder, M. (2003). Effects of declarative information in instructions for use. *IEEE International Professional Communication*, 29-33.

Karreman, J., & Steehouder, M. (2004). Some effects of system information in instructions for use. *IEEE Transactions on Professional Communication, 47*(1), 34-43.

Karreman, J., Ummelen, N., & Steehouder, M. (2005). Procedural and declarative information in user instructions: What we do and don't know about these information types. *IEEE International Professional Communcation Conference Proceedings*, 328-333.

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River, New Jersey: Pearson Prentice Hall.

Kester, L., Kirschner, P. A., & Van Merrienboer, J. J. (2006). Just-in-time information presentation: Improving learning a troubleshooting skill. *Contemporary Educational Psychology, 31*(2), 167-185.

Kester, L., Kirschner, P. A., Van Merrienboer, J. J., & Baumer, A. (2001). Just-in-time information presentation and the acquisition of complex cognitive skills. *Computers in Human Behavior, 17*, 373-391.

Kieras, D. E., & Bovair, S. (1984). The role of a mental model in learning to operate a device. *Cognitive Science, 8*(3), 255-273.

Kirkland, M. R., & Saunders, M. A. P. (1991). Maximizing student performance in summary writing: Managing cognitive load. *TESOL Quarterly, 25*(1), 105-121.

Kolodner, J. L., Owensby, J. N., & Guzdial, M. (2004). Case-based learning aids. In D. H. Jonassen (Ed.), *Handbook of Research on Educational Communications and Technology* (pp. 829-861). Mahwah, NJ: Lawrence Erlbaum Associates.

Konoske, P. J., & Ellis, J. A. (1991). Cognitive factors in learning and retention of procedural tasks. In R. F. Dillon & J. W. Pellegrino (Eds.), *Instruction: Theoretical and Applied Perspectives* (pp. 47-70). New York: Praeger.

Kontogiannis, T., & Sheperd, A. (1999). Training conditions and strategic aspects of skill transfer in a simulated process control task. *Human-Computer Interaction, 14*, 355-393.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Langan-Fox, J., Platania-Phung, C., & Waycott, J. (2006). Effects of advance organizers, mental models and abilities on task and recall performance using a mobile phone network. *Applied Cognitive Psychology, 20*, 1143-1165.

Langan-Fox, J., Waycott, J., & Albert, K. (2000). Linear and graphic advance organizers: Properties and processing. *International Journal of Cognitive Ergonomics, 4*(1), 19-34.

Lannon, J. M. (2008). *Technical Communication*. New York: Pearson-Longman.

Lewis-Beck, M. S. (1980). *Applied regression: An introduction*. Newbury Park, CA: Sage Publications.

Mayer, R. E. (1981). The psychology of how novices learn computer programming. *ACM Computing Surveys*, *13*(1), 121-141.

Mayer, R. E. (1989). Models for understanding. *Review of Educational Research*, *59*(1), 43-64.

Mayer, R. E. (2001). *Multimedia learning*. Cambridge, UK: Cambridge University Press.

Mayer, R. E. (2003). *Learning and Instruction*. Upper Saddle River, NJ: Merrill Prentice Hall.

Mayer, R. E., & Bromage, B. K. (1980). Different recall protocols for technical texts due to advance organizers. *Journal of Educational Psychology, 72*(2), 209-225.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30-46.

Morris, N. M., & Rouse, W. B. (1985). The effects of type of knowledge upon human problem solving in a process control task. *IEEE Transactions on Systems, Man, & Cybernetics, 15*(6), 698-707.

Norman, D. A. (1987). Some observations on mental models. In R. M. Baecker & W. A. S. Buxton (Eds.), *Readings in Human-Computer Interaction: A Multidisciplinary Approach* (pp. 241-244). Los Altos, CA: Morgan Kaufmann.

Novick, L. R., & Bassok, M. (2005). Problem solving. In K. J. Holyoak (Ed.), *The Cambridge handbook of thinking and reasoning* (pp. 321-349). Cambridge: Cambridge University Press.

Nunnally, J. C. (1967). *Psychometric theory*. New York: McGrawy-Hill.

Paas, F., Camp, G., & Rikers, R. (2001). Instructional compensation for age-related cognitive declines: Effects of goal specificity in maze learning. *Journal of Educational Psychology, 93*(1), 181-186.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1-4.

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. A. (2009). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest, 9*(3), 105-119.

Patrick, J. U., & Haines, B. (1988). Training and transfer of fault-finding skills. *Ergonomics, 31*(2), 193-210.

Payne, S. J., Howes, A., & Hill, E. (1992). Conceptual instructions derived from an analysis of device models. *International Journal of Human-Computer Interaction, 4*(1), 35-57.

Pirolli, P., & Recker, M. M. (1994). Learning strategies and transfer in the domain of programming. *Cognition and Instruction, 12*(3), 235-275.

Pressley, M. (2006). *Reading instruction that works: The case for balanced teaching*. New York: Guilford.

Quintana, C., Carra, A., Krajcik, J., & Soloway, E. (2002). Learner-centered design: Reflections and new directions. In J. M. Carroll (Ed.), *Human-Computer Interaction in the New Millennium* (pp. 605-626). New York: Addison-Wesley.

Reder, L. M., Charney, D. H., & Morgan, K. I. (1986). The role of elaborations in learning a skill from an instructional text. *Memory & Cognition, 14*(1), 64-78.

Reigeluth, C. M. (1983). Meaningfulness and instruction: Relating what is being learned to what a student knows. *Instructional Science, 12*, 197-218.

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist, 38*(1), 15-22.

Roediger, H. L., & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implication for educational practice. *Perspectives on Psychological Science, 1*(3), 181 - 210.

Roediger, H. L., & Karpicke, J. D. (2006b). Test enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249 - 255.

Rubio, S., Diaz, E., Martin, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology: An International Review, 53*(1), 61-86.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*(4), 207-217.

Schneider, W. (1985). Training high-performance skills: Fallacies and guidelines. *Human Factors, 27*(3), 285-300.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420-428.

Simon, H. A. (1969). *The sciences of the artificial.* Cambridge, MA: MIT Press.

Simutech. (2009). Troubleshooting Electrical Circuits (Version 4) [Computer Software]. Ottawa, ON: Simutech Multimedial Inc.

Smith, E. E., & Goodman, L. (1984). Understanding written instructions: The role of an explanatory schema. *Cognition and Instruction, 1*(4), 359-396.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257-285.

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General, 119*(2), 176-192.

van Belle, G. (2008, September 4). *Statistical rules of thumb*. Downloaded from www.vanbelle.org. Retrieved 5/24, 2011

Van Der Meij, H. (2003). Minimalism revisited. *Document design, 4*(3), 212-233.

Van Der Meij, H., & Lazonder, A. (1993). Assessment of the minimalist approach to computer user documentation. *Interacting with Computers, 5*(4), 355-370.

Van Gog, T., & Kester, L. (2011). *A test of the testing effect: Acquiring problem-solving skills from worked examples*.Unpublished manuscript.

Van Merrienboer, J. J., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist, 38*(1), 5-13.

Waern, Y. (1993). Varieties of learning to use computer tools. *Computers in Human Behavior, 9*(2), 323-339.

Wiedenbeck, S., & Zila, P. L. (1997). Hands-on practice in learning to use software: A comparison of exercise, exploration, and combined formats. *ACM Transactions on Computer-Human Interaction*, *4*(2), 169-196.

Wittwer, J., & Renkl, A. (2008). Why instructional explanations often do not work: A framework for understanding the effectiveness of instructional explanations. *Educational Psychologist, 43*(1), 49-64.

VITA

ELSA EIRIKSDOTTIR

Eiriksdottir was born in Reykjavik, Iceland. She received a B.A. in Psychology from University of Iceland, Reykjavik, Iceland in 1999 and after receiving a Fulbright Scholarship relocated to Atlanta to pursue a graduate degree in Engineering Psychology at Georgia Institute of Technology. Elsa received a M.S. in Psychology from Georgia Institute of Technology, Atlanta, Georgia in 2007.